NONPARAMETRIC ESTIMATION OF A MULTIFACTOR HEATH-JARROW MORTON MODEL: AN INTEGRATED APPROACH

BY

**ANDREW JEFFREY, DENNIS KRISTENSEN,
OLIVER LINTON, THONG NGUYEN
and PETER C. B. PHILLIPS**

COWLES FOUNDATION PAPER NO. 1143

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281**

2006

http://cowles.econ.yale.edu/

Nonparametric Estimation of a Multifactor Heath-Jarrow-Morton Model: An Integrated Approach

ANDREW JEFFREY
Yale School of Management

DENNIS KRISTENSEN
London School of Economics

OLIVER LINTON
London School of Economics

THONG NGUYEN
Alpha Simplex Group

PETER C. B. PHILLIPS
Yale University

ABSTRACT

We propose a new nonparametric estimator for the volatility structure of the zero-coupon yield curve inside the Heath-Jarrow-Morton framework. The estimator incorporates cross-sectional restrictions along the maturity dimension, and also allows for measurement errors, which can arise from estimation of the yield curve from noisy data. The estimates are implemented with daily CRSP bond data.

KEYWORDS: continuous-time estimation, dynamic panel data model, Heath-Jarrow-Morton model, measurement errors, nonparametric

A well-known limitation of one-factor models, regardless of whether they belong to the more traditional Markovian framework typified by Vasicek (1977) and Cox, Ingersoll, and Ross (1985) or the more recent non-Markovian approach pioneered by Ho and Lee (1986) and Heath, Jarrow, and Morton (1992) (HJM hereafter), is that they poorly capture empirical dynamics either of short rates (the most commonly used state variable) or the entire term structure. This limitation has

We would like to thank Eric Renault, Federico Bandi, and a referee for helpful comments and Christian Huse for research assistance. We are grateful to the National Science Foundation and the Financial Markets Group of the London School of Economics for financial support. Address correspondence to Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK, or e-mail: o.linton@lse.ac.uk.

Journal of Financial Econometrics, Vol. 2, No. 2, pp. 251–289
 DOI: 10.1093/jjfinec/nbh010

motivated the study of multifactor models, especially in practical work. Two factor models with a rich variety of second-factor specifications have been examined in the literature. The short-term interest rate is usually the first factor, while the long rate, the spread between long rates and short rates, inflation, central tendency (the long-term mean of short-term interest rates), and volatility of short-term interest rates are frequently chosen second factors. Examples of two-factor models can be found in Brennan and Schwartz (1979), Schaefer and Schwartz (1984), Litterman and Scheinkman (1991), Chen and Scott (1992), Longstaff and Schwartz (1992), Pearson and Sun (1994), Duffie and Singleton (1997), and Knight, Li, and Yuan (1999) among others. More flexible three-factor models and beyond have been studied in, for example, Chen and Scott (1993), Balduzzi et al. (1996), Duffie and Kan (1996), and Boudoukh et al. (1997).

Recent theoretical and empirical studies of non-Markovian models inside the HJM framework have also favored the multifactor specification to enable richer dynamics [see, e.g., Bliss and Ritchken (1996), Inui and Kijima (1998), Bühler et al. (1999), de Jong and Santa-Clara (1999), and Pearson and Zhou (2000)]. Moving away from the simple one-factor framework introduces extra analytical complexity however. Consequently, out of mathematical convenience, most models so far studied are affine. That is, they have drift and diffusion functions defined as linear functions of the state variable(s), as proposed in Vasicek (1977) and typified in the work of Cox, Ingersoll, and Ross (1985). The affine structure by itself imposes heavy restrictions on the specified dynamics, besides the fact that most affine models impose additional restrictions, thereby falling short of being "maximal" in the terminology of Dai and Singleton (2000).

We propose a nonparametric multifactor HJM model. The model has two main advantages compared to most of the existing literature: First, the nonparametric specification allows the functional forms of interest to be minimally restricted. We also allow the volatility structures associated with each Brownian motion to depend on the entire set of specified state variables, departing from a common practice of associating each Brownian motion with a separate state variable. These features lead to a model that encompasses most HJM models proposed so far in the literature. Second, our model directly incorporates measurement errors when the dynamics of the yield curves are initially specified. This is in contrast to most of the empirical work on yield curves, where measurement errors often are added as an afterthought.

In the econometric part, we set up a fully nonparametric estimator of the diffusion term of the yields and derive its asymptotic properties. To the best of our knowledge, only two studies have applied nonparametric techniques to estimate the volatility structure in the HJM setting: Pearson and Zhou (1999) and Jeffrey, Linton, and Nguyen (1999a). However, the former does not allow the volatility structure to depend on interest rate levels, does not allow for observation errors, nor provides any asymptotic results for the proposed estimator. The latter only considers single-factor models and does not allow for observation errors. So our approach can be seen as a generalization of these studies. The estimation method is based on a combination of some recent advances in the nonparametric literature.

Along the maturity dimension, we draw on the approach put forward by Linton et al. (2001) (hereafter LMNT), while information from the evolution of the observed interest rates over time is used in a similar fashion to the approach taken in the literature on nonparametric kernel estimation of diffusion processes [see, e.g., Florens-Zmirou (1993), Jiang and Knight (1997), Stanton (1997), and Bandi and Phillips (2003)]. All these studies rely on so-called in-fill asymptotics, where the time distance between the observations shrink to zero; this allows us to reconstruct the sample path in the limit and from that extract enough information to obtain an estimator of the diffusion function as the instantaneous conditional variance. For a fixed-time distance between observations, our estimator will, however, be biased. But so far, very few non/semiparametric estimation methods for diffusion processes based on a fixed-time distance have been proposed in the literature — one example is the seminal work by Aït-Sahalia (1996) — and certainly not for HJM-type models. So here we shall rely on in-fill estimators despite this drawback. For a relatively high sample frequency over time, we believe that our proposed method is able to give a fairly accurate estimate of the diffusion term.

Even without the presence of observation errors, the proposed nonparametric estimator is of interest since it has a better convergence rate than the ones proposed by Pearson and Zhou (1999) and Jeffrey, Linton, and Nguyen (1999a). One can consider the HJM model as a dynamic panel data model, where yields are observed both over time and across maturities. Our estimation procedure then incorporates the full information set, that is, observed yields across all maturities and over time, when estimating the volatility at a specific maturity point. In contrast, the estimators of Pearson and Zhou (1999) and Jeffrey, Linton, and Nguyen (1999a) only use the yields observed at the maturity point of interest. Thus, we utilize all available information in the estimation, while the two other studies only use the information at the maturity point of interest. This again means that our estimator will be more precise. This is of particular interest when data are only available over a short time horizon and in the estimation of multifactor models (where nonparametric estimators will suffer from the well-known "curse of dimensionality").

The plan of the paper is as follows. In Section 1 the prototype one-factor model is specified. In Section 2 we derive our estimator. In Section 3 we present its asymptotic properties. We extend the methodology to cover the multifactor case in Section 4. Section 5 is devoted to the empirical estimation of a one-factor and two-factor model using U.S. bond data. Section 6 concludes. All proofs are given in the appendix.

1 SINGLE-FACTOR NONPARAMETRIC HJM MODEL

Let $P_t\ (\tau)$ denote the price of a one dollar face value, default-free, zero-coupon bond at time t with time to maturity τ; that is, it will mature at time $t+\tau$. The instantaneous forward rate at time t for date $t+\tau$, denoted $f_t(\tau)$, is defined by $f_t(\tau) = -\partial \ln(P_t(\tau))/\partial\tau$ and the yield at time t with maturity date $t+\tau$, denoted $y_t(\tau)$, is defined in terms of forward rates by $y_t(\tau) = -\tau^{-1}\int_0^\tau f_t(s)\,ds$. In the original HJM framework, the term structure is modeled in terms of forward rates. However, as already argued in Jeffrey, Linton, and Nguyen (1999a), determining forward rates in practice via curve-fitting procedures often proves to be sensitive to the method adopted. Estimation of yields on the other hand is typically less sensitive to the method used, intuitively because yields are averages of forward rates. Consequently we choose to portray the term structure using yields. However, the estimator developed here can be adapted to forward curve evolution with minor modifications.

Within the original HJM framework, a single error process introduces uncertainty into the bond market. The uncertain evolution of each yield with fixed time to maturity τ is in this setting characterized by the following stochastic differential equation (SDE),

$$dy_t(\tau) = \alpha(\omega_t, t, \tau)dt + \gamma(\omega_t, t, \tau)dW_t^*, \tag{1}$$

where $\{W_t^*\}$ is a one-dimensional standard Brownian motion, while ω_t indicates the possible dependence of the term structure on various variables, including itself, observed up to time t. More specifically, $\omega_t \in \mathcal{F}_t$, where $\mathcal{F}_t$ indicates all available information just before time t. We shall restrict our attention to drift and volatility functions of the form

$$\alpha(\omega_t, t, \tau) = \alpha(\omega_t, \tau), \quad \gamma(\omega_t, t, \tau) = \gamma(X_t, \tau),$$

where $\{X_t\}$ is some underlying factor. This class of models is very general, covering most of the HJM specifications proposed thus far in the literature. We shall not impose any parametric restrictions on either of the two functions, instead we shall rely on nonparametric estimators of them. The asymptotic properties of the estimator are derived imposing only weak restrictions on the functional form. The state variable X can be chosen in various ways, but we have to require that it is observable. In our empirical study, we shall choose X as a proxy of the spot rate, but any observed variable which the practitioner believes to be the driving factor of the volatility term of the yield curve can be used.

As shown in the original HJM article, the no-arbitrage restriction dictates that the drift function of the forward curve evolution is just a function of the volatility structure if we are in a risk-neutral world. For the yield curve evolution, a similar restriction on the drift function can be shown to hold. Specifically,

$$\alpha(\omega_t, \tau) = \frac{\partial y_t(\tau)}{\partial \tau} + \frac{y_t(\tau) - r_t}{\tau} + \frac{1}{2}\tau\gamma^2(X_t, \tau),$$

where $r_t \equiv f_t(t)$ denotes the instantaneous interest rate [see Jeffrey, Linton, and Nguyen (1999a) for a derivation of the above equation]. Together with the knowledge of the market price of risk, the dynamics of the yield curve under the real-world measure can be recovered. A common objective is to use the above dynamics to price fixed-income instruments, and risk-neutral martingale pricing only requires the specification of the dynamics in a risk-neutral world. In this situation we therefore only need to know γ, from which we can back out α (it is not necessary to know the market price of risk in this world). In conclusion, for pricing based on an HJM model such as the one considered here, we only need to estimate the volatility structure.

Given discrete observations of the process $\{y_t(\tau)\}$ from only one fixed maturity τ, we can obtain an estimate of the volatility function $\gamma(x, \tau)$, as shown in Jeffrey, Linton, and Nguyen (1999a). Their estimator was

$$\tilde{\gamma}^2(x,\tau) = \frac{1}{\Delta}\frac{\sum_{i=1}^n K_{h_x}(X_{t_i} - x)(y_{t_{i+1}}(\tau) - y_{t_i}(\tau))^2}{\sum_{i=1}^n K_{h_x}(X_{t_i} - x)}, \tag{2}$$

where $K_h(\cdot) = K(\cdot/h)/h$ and K is a kernel, while h_x is a bandwidth sequence. Under various regularity conditions (including that $h_x \to 0$ and $nh_x \to \infty$), they showed

$$\sqrt{\frac{h_x\bar{L}(T,x)}{\Delta t}}[\tilde{\gamma}^2(x,\tau) - \gamma^2(x,\tau)] \xrightarrow{d} N\left(0, 4\|K\|_2^2\gamma^4(x,\tau)\right)$$

as $n \to \infty$, where $\|K\|_2^2 = \int_{-\infty}^{+\infty} K^2(z)\,dz$, while $\bar{L}(T, x)$ is the chronological time of the process $\{X_t\}$, a concept that will be discussed in more detail later. The estimator can be seen as an extension to HJM models of the approach pioneered by Jiang and Knight (1997) and Stanton (1997) for Markovian interest rate models.

In the original HJM models, a finite set of Brownian motions serves as the source of randomness in the economy and drives the dynamics of the yield curves. This construction helps to make the market complete, but at the cost of inducing stochastic singularity into the model: There exists a linear combination of points along the yield curves which is deterministic, that is, without the presence of the Brownian motion shocks.[1]

In the following, we will introduce a new, general model that is highly flexible and removes stochastic singularity, while retaining tractability and allowing for estimation of γ. We will assume that the specification of Equation (1) holds for some true yields $\{y_t^*(\tau)\}$, but that the observed yields are given by

$$y_{t_i}(\tau_j) = y_{t_i}^*(\tau_j) + z_{t_i}(\tau_j), \quad i = 1, \ldots, n, \quad j = 1, \ldots, J, \tag{3}$$

where the $z_{t_i}(\tau_j)$'s are a sequence of random variables that can be seen as observation errors. We assume that the underlying process, $\{y_t^*(\tau)\}$, solves the same SDE as before, namely

$$dy_t^*(\tau) = \alpha(\omega_t, \tau)dt + \gamma(X_t, \tau)dW_t^*. \tag{4}$$

The driving, one-dimensional factor, $\{X_t\}$, is observed without measurement errors and also solves an SDE,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t. \tag{5}$$

[1] For example, in the one-factor model above, we can have the following deterministic relationship for any two yield dynamics: $dy_t(\tau_1)\gamma(\omega_t, t, \tau_2) - \alpha(\omega_t, t, \tau_1)\gamma(\omega_t, t, \tau_2)dt = dy_t(\tau_2)\gamma(\omega_t, t, \tau_1) - \alpha(\omega_t, t, \tau_2)\gamma(\omega_t, t, \tau_1)dt$.

We do not impose any restrictions on the dependence between the two Brownian motions $\{W_t\}$ and $\{W_t^*\}$, which may be identical.

Our model covers all one-factor HJM models that have been employed in the past literature, including the one-factor models in Jeffrey, Linton, and Nguyen (1999a) and Pearson and Zhou (1999).[2] In contrast to more radical approaches such as that of Kennedy (1994, 1997), where an infinite-dimensional source of shocks drives the yield curve, the model proposed above can be considered as a slightly generalized version of the original HJM model. The model aims to incorporate measurement or observation errors while preserving as much as possible of the shock structure of the original HJM-model. In particular, when $\{W_t\} = \{W_t^*\}$, the common Brownian motion(s) that drives the yield curve dynamics still provides the only source of economic uncertainty embedded in the underlying model, while the idiosyncratic shocks added to each point of the yield curves are statistical noise. The underlying economy is therefore still a complete market, where any instrument is hedgeable, while the observed economy is not necessarily complete anymore. We thus take a halfway approach between the original HJM models and that of Kennedy (1994, 1997) and Santa-Clara and Sornette (2001).

The errors, $\{z_t(\tau)\}$, can arise from a variety of sources, a notable one being that the available yield data, $\{y_{t_i}(\tau_j)\}$, may have been obtained from some preliminary yield curve fitting procedure such as splines [McCulloch (1971, 1975)], bootstrapping [Fama and Bliss (1987)], or kernel smoothing (LMNT).[3] We assume that $z_t(\tau)$ solves

$$dz_t(\tau) = m(\omega_t, \tau)\,dt + \nu(X_t, \tau)\,dW_t(\tau) \tag{6}$$

for each τ, where $\{W_t(\tau): t \geq 0, \tau \geq 0\}$ is a stochastic process such that for each fixed value of τ, $\{W_t(\tau): t \geq 0\}$ is a Brownian motion. We shall allow for correlation between the errors at different observed maturities, $z_t(s)$ and $z_t(\tau)$, by assuming that $E[W_{t_1}(s)\ W_{t_2}(\tau)] = (t_1 \wedge t_2)\rho(|s-\tau|)$, where $\tau \mapsto \rho\ (\tau)$, $\tau \geq 0$ is a decreasing function satisfying $\rho\ (0) = 1$. For the LMNT kernel method, it can be expected that $\mathrm{cov}(z_t(s), z_t(\tau)) = 0$ for any s, τ with $|s-\tau|$ larger than the bandwidth used in preliminary estimation of the yields, as is the case for standard kernel regression or density estimators; our specification includes this important case. We furthermore assume that $\{W_t(\tau)\}$ is independent of $\{W_t^*\}$, such that the underlying yield structure $\{y_t^*(\tau)\}$ and the observation errors $\{z_t(\tau)\}$ are uncorrelated.

The above strategy of adding measurement errors to the yield curve to remove stochastic singularities is found elsewhere in the literature since such

[2] The two-factor model of Pearson and Zhou (2000) is essentially covered by the multi-factor model proposed in the next section.

[3] Zero-coupon yield curves are not observable with traded instruments in the fixed-income markets because bonds with time to maturity greater than one year are normally coupon-bearing bonds. So the (zero coupon) yield curves must be extracted by yield curve fitting procedures [see e.g., McCulloch (1971, 1975)]. Moreover, the data used to carry out these procedures do not correspond exactly to the theoretical price. Specifically, we usually observe quotes of bid and ask prices obtained from a telephone survey of the registered bond dealers. Also, there are tax difference and liquidity effects that can be interpreted as providing random errors in the observed bond prices.

singularities are often present in both Markovian and non-Markovian interest rate models. In Markov models, it is often the case that structural parameters controlling the dynamics of the short-term interest rates (the often-used state variables) are estimated, with the (implicit) assumptions that the short rates are observed without errors. Then, to reconcile the fact that other observed yields with maturity longer than zero are not perfectly correlated, which they should be due to being driven by the same state variables, it is necessary to introduce observation errors for other yields. So in effect we would "shift" all observation errors toward other yield levels, and make a (perhaps unrealistic or unreasonable) assumption that we could observe the short rate with utmost accuracy. In the HJM framework, there is also an inevitable need to introduce observation errors if we are to reconcile for the less than perfect correlation observed in yield data across the maturity dimension. One possible approach, which is somewhat comparable to the common approach in the Markovian literature [see Dai and Singleton (2000), among others], is to assume that we can observe a carefully chosen number, N, of points along the yield curves without measurement errors. For instance, for a two-factor HJM model, one may assume that we can observe yields at three different maturities without measurement errors. This approach is similar to that of the Markovian literature, in the sense that we assume that there are some "special" points that are observed without errors, and "shift" all observation errors to other points. The approach is, however, not very appealing for a number of reasons, one being that there is no natural candidate for these N points. Our model offers a more realistic setup where measurement errors appear at all maturities.

2 ESTIMATION METHOD

In this section we introduce our estimation technique based on observations of the yields and the underlying factor over time. The introduction of observation errors makes the estimation issue much harder to deal with than in Jeffrey, Linton, and Nguyen (1999a) since the presence of the errors lead to a bias term. But by combining observed yields across different maturities, it proves possible to kill this bias term asymptotically.

As already noted, we only observe the yield curve and the underlying factor a discrete number of times n, $0 < t_1 < t_2 < \cdots < t_n = T$. Let $\Delta y_i(\tau) = y_{t_{i+1}}(\tau) - y_{t_i}(\tau)$ and $X_i = X_{t_i}$. For simplicity it is assumed that all time intervals t_i to t_{i+1} are equally spaced, that is, $t_{i+1} - t_i \equiv \Delta$ for all i and consequently $T = n\Delta$. At each time point, t_i, we observe a crosssection of yields across different maturities, $\tau_1, \ldots, \tau_J$. So the dataset that our estimation method will based on is given as $\{X_{t_i}, y_{t_i}(\tau_j) : i = 1, \ldots, n, j = 1, \ldots, J\}$.

Given the mutual independence between $\{W_t^*\}$ and $\{W_t(\tau)\}$ together with the infinitesimal characteristics of $\{y_t^*(\tau)\}$ and $\{z_t(\tau)\}$, it is seen that

$$E[\Delta y_t(s)\Delta y_t(\tau) \mid \mathcal{F}_t] \simeq \gamma(X_t, s)\gamma(X_t, \tau)\Delta + \nu(X_t, s)\nu(X_t, \tau)\rho(|s - \tau|)\Delta, \tag{7}$$

as $\Delta \to 0$. We will use Equation (7) as an estimating equation upon which we build our estimator. But from this formula it is evident that we cannot distinguish between the variance of the observation errors and the one of $y_t^*(\tau)$ if we only combine observables from the same maturity. The presence of the second term on the right-hand side of Equation (7) may potentially lead to biased nonparametric estimates of γ since one may not be able to disentangle the first term from the second. This is particularly a problem in a nonparametric framework; for example, the naïve estimator $\tilde{\gamma}^2(x, \tau)$ in Equation (2) will be centered around $\gamma^2(x, \tau) + \nu^2(x, \tau)$, hence biased, in the presence of observation errors. However, a well-known result from the kernel regression literature [see, e.g., Robinson (1997)] states that a kernel estimator of a regression function evaluated at two different points will be mutually asymptotically independent even in the presence of time-series dependence in the errors. We take advantage of this property and generate estimating equations that only include cross-products at different maturity points, and thereby kill the asymptotic bias induced by the correlation between the observation errors. In doing so, we are able to extract the unknown volatility function γ.

So by only including cross-products of yields observed at different maturities, we may ignore the second term on the right-hand side. This leaves us with what is really a multiplicative nonparametric regression model structure. That is, the left-hand side of Equation (7) can be substituted with the observed yields, but the right-hand side is effectively the product of two unknown functions. It is not immediately obvious how to extract an estimate of γ here in a nonparametric setting [in the parametric case, standard generalized methods of moments (GMM) would work]. Similar estimation problems have been analyzed in Newey and Powell (2003), Mammen, Linton, and Nielsen (1999), Darolles, Florens, and Renault (2001), and in particular Breiman (1991); we refer to Carrasco, Florens, and Renault (2003) for a review. We shall import the technology of those studies to solve this problem.

We first define our criterion function. In order to deal with the correlation between the observation errors, we only include cross-products from different maturities. Leaving out products of observations from the same maturity will enable us to control this observation bias under suitable conditions. Contrary to the approach in the related study by Jeffrey, Linton, and Nguyen (1999a), where only information along the diagonal of the moment condition matrix is used, we here use only information from the off-diagonal elements.[4] The resulting volatility structure estimate may then be used together with the diagonal information afterwards to back out and make inference about ν and ρ. Our criterion function

[4] Our asymptotic results will in fact go through even if we did include the diagonal terms in the estimation. This owes to the fact that the asymptotics are based on $J \to \infty$, such that the diagonal terms will be asymptotically negligible relative to the off-diagonal ones. In a finite sample, however, leaving out the diagonal terms will most likely improve the quality of the estimator.

is basically a localized sum of the squared residuals of Equation (7),

$$\int \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} K_{h_x}(X_i - x) K_{h_\tau}(\tau_j - \tau) K_{h_\tau}(\tau_k - s)[\Delta y_i(\tau_j)\Delta y_i(\tau_k) - \gamma(x,s)\gamma(x,\tau)\Delta]^2 d(x,s,\tau),$$

which is to be minimized with respect to the function γ; this yields our estimator, $\hat{\gamma}$. Here, h_τ and h_x are two bandwidth sequences, while K still denotes a kernel. Using (functional) differentiation of the above expression with respect to γ as in LMNT, we obtain the first-order condition that $\hat{\gamma}$ must satisfy,

$$0 = \frac{1}{J^2} \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} K_{h_x}(X_i - x)\Delta y_i(\tau_j)\Delta y_i(\tau_k) K_{h_\tau}(\tau_j - \tau) \int \hat{\gamma}(x,s) K_{h_\tau}(\tau_k - s)\, ds$$
$$- \frac{1}{J^2} \sum_{i=1}^{n-1} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} K_{h_x}(X_i - x)\hat{\gamma}(x,\tau) K_{h_\tau}(\tau_j - \tau) \int \hat{\gamma}^2(x,s) K_{h_\tau}(\tau_k - s)\, ds\Delta.$$

The function $\hat{\gamma} \equiv 0$ is always a solution to Equation (9). Also, if $\hat{\gamma}$ is a solution so is $-\hat{\gamma}$. So in order to obtain a nontrivial unique estimator, we restrict the class of feasible solutions to satisfy $\gamma > 0$. The resulting solution $\hat{\gamma}$ can also be written as

$$\hat{\gamma}(x,\tau) = \frac{\int \hat{H}_1(x, s, \tau)\hat{\gamma}(x, s) ds}{\int \hat{H}_2(s, \tau)\hat{\gamma}^2(x, s) ds}, \tag{9}$$

where

$$\hat{H}_1(x,s,\tau) = \frac{1}{J^2} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} K_{h_\tau}(\tau_j - \tau) K_{h_\tau}(\tau_k - s)\hat{v}(x, \tau_j, \tau_k), \tag{10}$$

with

$$\hat{v}(x,s,\tau) = \frac{\frac{1}{\Delta}\sum_{i=1}^{n} K_{h_x}(X_i - x)\Delta y_i(s)\Delta y_i(\tau)}{\sum_{i=1}^{n} K_{h_x}(X_i - x)}, \tag{11}$$

and

$$\hat{H}_2(s,\tau) = \frac{1}{J^2} \sum_{j=1}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} K_{h_\tau}(\tau_j - \tau) K_{h_\tau}(\tau_k - s). \tag{12}$$

Equation (9) is a nonlinear integral equation involving linear operators defined by $\hat{H}_1$ and $\hat{H}_2$. These two quantities depend only on observed data and indeed are just kernel weighted sample averages.

Equation (9) suggests the following iterative procedure for the calculation of $\hat{\gamma}(x, \cdot)$: For any x in the domain of $\{X_t\}$, calculate

$$\hat{\gamma}^{[m]}(x,\tau) = \frac{\int \hat{H}_1(x,s,\tau)\hat{\gamma}^{[m-1]}(x,s)\, ds}{\int \hat{H}_2(s,\tau)\hat{\gamma}^{[m-1]}(x,s)^2\, ds}, \quad m = 1,2,\ldots,$$

for some given starting value $\hat{\gamma}^{[0]}(x,\cdot) = \gamma^*(x,\cdot)$, an obvious choice being $\gamma^*(x,\cdot) = \tilde{\gamma}(x,\cdot)$ as given in Equation (2). As J and $n \to \infty$, $\hat{H}_1(x,s,\tau)$ and $\hat{H}_2(x,s,\tau)$ converge toward positive functions in probability, implying that $\hat{\gamma}^{[m]}(x,\tau)$ will be positive with probability one if $\gamma^*(x,\tau)$ has been chosen to be positive. The above integrals are one dimensional and can be computed numerically. This iterative method is called successive approximation. For a detailed discussion we refer the reader to Kantorovich and Akilov (1982) and Luenberger (1969); see also Hastie and Tibshirani (1990), Mammen, Linton, and Nielsen (1999), and LMNT for related computations. Breiman (1991) provides some regularity conditions under which this algorithm converges in a related multiplicative nonparametric regression model. But as stated there, what is difficult is not convergence, but the possible presence of local minima to either of which the algorithm may converge.

3 ASYMPTOTIC PROPERTIES IN THE SINGLE-FACTOR CASE

Before presenting the asymptotic properties of the estimator defined above, we first introduce some additional concepts and assumptions.

To simplify our setup, we normalize the maturities to lie in the unit interval $[0,1]$, and furthermore assume that we are able to observe the yields at maturities over an increasingly fine grid, $\tau_j = j/J$, $j = 1,\ldots,J$, where $J \to \infty$. To kill the bias term in Equation (7) we need to restrict the correlation structure of the observation errors across different maturities. The correlation between $z_t(\tau_j)$ and $z_t(\tau_k)$ is determined by the one between $W_t(\tau_j)$ and $W_t(\tau_k)$. We then impose the following restriction on the correlation structure across observed maturities,

$$E[W_t(\tau_j)W_t(\tau_k)] = t\rho_J(|\tau_j - \tau_k|) = t\rho(|j-k|), \tag{13}$$

where $j \mapsto \rho(j)$, $j \geq 0$ is a decreasing function satisfying $\rho(0) = 1$ and $\lim_{j\to\infty} \rho(j) = 0$.[5]

This setup can be interpreted as a case of increasing-domain spatial asymptotics [cf. Cressie (1993:100)], where the range of unnormalized observed maturities grow as $J \to \infty$. This will yield the type of correlation function in Equation (13) for the normalized ones. To see this, let $\tilde{\tau}_1 < \tilde{\tau}_2 < \cdots < \tilde{\tau}_J$ denote the unnormalized ordered maturities at which we observe yield. Assume that they are equidistant and we observe yields over an increasing range of maturities such that $\tilde{\tau}j = \delta j$ for some $\delta > 0$. Moreover, assume that $E[W_t(\tilde{\tau}_j)W_t(\tilde{\tau}_k)] = t\rho(|\tilde{\tau}_j - \tilde{\tau}_k|)$. The normalized maturities are then given as $\tau_j = \tilde{\tau}_j/\tilde{\tau}_J = j/J$, and

$$\rho(|\tilde{\tau}_j - \tilde{\tau}_k|) = \rho(\tilde{\tau}_J|\tau_j - \tau_k|) = \rho(\delta J|\tau_j - \tau_k|) = \rho(\delta|j-k|).$$

So under the assumption of observing maturities over an increasing domain, the correlation function in terms of normalized maturities takes the desired form.

[5] An alternative way to ensure that the observation errors do not interfere with the asymptotic properties of the estimator is to assume that they disappear asymptotically. That is, $m = m_{nJ}$ and $\nu = \nu_{nJ}$ both go to zero as $n, J \to \infty$. In some cases this might be seen as a more realistic setup. We thank an anonymous referee for pointing this out to us.

In conclusion, if we have observed yields across a sufficiently large domain, Equation (13) is not an unreasonable assumption. Observe that while, for the unnormalized maturities, the correlation structure across maturities remains unaltered as $J \to \infty$, for any two fixed normalized maturities, $s, \tau \in [0,1]$, $E[W_t(s)\ W_t(\tau)] = t\rho\ (|s-\tau|\ J) \to 0$, as $J \to \infty$, so that in the limit the correlation between the observation errors at any two distinct normalized maturities is zero. In any finite sample of maturities, however, there may be limited correlation between any two normalized maturities.

This way of restricting the correlation structure of the errors is very much standard in the literature on nonparametric kernel regression with correlated/time-series errors [see, e.g., Robinson (1997) and Opsomer, Wang, and Yang (2001)]. Below we impose restrictions on the rate of decay with which $\rho(j)$ goes to zero as $j \to \infty$. When the correlation function decreases at order j^{-a} with $0 < a \leq 1$, we have so-called long-range dependence, and for $a > 1$, short range. Here we shall assume short-range dependence. This assumption combined with the employment of higher-order (bias-reducing) kernels[6] will allow us to disregard the correlation across maturities, and obtain the standard nonparametric regression convergence rate, $\sqrt{Jh_\tau}$, along the maturity dimension.

If the observation errors exhibit long-range dependence, we conjecture that our estimator will still be consistent, but will converge with decreased rate along the maturity dimension. Robinson (1997:Theorem 1) shows in a simple setting how the convergence rate of a kernel regression estimator decreases as a decreases. His results for the long-range dependence case are nontrivial, however, and cannot be directly translated to our more complicated setup. We will therefore not give any theoretical results for this case.

If the correlation function ρ does not satisfy Equation (13) for the data at hand, our proposed estimator will be biased asymptotically. We show this at the end of this section.

Next, we turn to the sampling over time. We do not wish to impose any stationarity assumptions on the factor $\{X_t\}$. Instead, we shall allow for (weak) nonstationarity, and rely on some recent results on nonparametric estimation of diffusion models by Bandi and Phillips (2003). In doing so, we need to introduce the important concepts of local and chronological time.

Definition 1 (chronological local time) *The chronological local time of the semimartingale $\{X_t\}$ at any point x in its domain over the time interval $[0,T]$ is defined as*

$$\bar{L}(T,x) = \frac{1}{\sigma^2(x)} L(T,x),$$

where $L(T,x) \equiv \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^T \mathbf{1}_{\{|X_t - x| < \varepsilon\}} \sigma(X_t)^2 dt$ is the local time of $\{X_t\}$.

[6] See Pagan and Ullah (1999; Section 2.7.2) for a discussion of such kernels.

For a detailed discussion of the above concepts, we refer to Revuz and Yor (1991) and Phillips and Park (1998). The following result can be found in Revuz and Yor (1991):

Lemma 2 (the occupation time formula) *For the semimartingale $\{X_t\}$ with quadratic variation process $\langle X\rangle_t$, and for every Borel function f of X_t,*

$$\int_0^T f(X_t)d\langle X\rangle_t = \int_{-\infty}^{+\infty} f(u)L(T,u)\,du. \tag{14}$$

Direct applications of the occupation time formula along with the definition of chronological local time provide the following two results, which will be used repeatedly in our proofs: For any Borel function f,

$$\int_0^T f(X_t)dt = \int_0^T \frac{f(X_t)}{\sigma^2(X_t)^2}d\langle X\rangle_t = \int_{-\infty}^{+\infty} f(u)\bar{L}(T,u)du$$

and further, for any kernel K and continuous function f,

$$\begin{aligned}\lim_{h\downarrow 0}\frac{1}{h}\int_0^T K\left(\frac{X_t-x}{h}\right)f(X_t)dt &= \lim_{h\downarrow 0}\frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{u-x}{h}\right)f(u)\bar{L}(T,u)\,du\\ &= \lim_{h\downarrow 0}\int_{-\infty}^{+\infty} K(v)f(x+hv)\bar{L}(T,x+hv)\,dv\\ &= \bar{L}(T,x)f(x)+o_P(1).\end{aligned}$$

A related concept is recurrence. We say that $\{X_t\}$ is recurrent, if at every point x on its support, $\bar{L}(T,x)\to\infty$ as $T\to\infty$. This basically implies that the process revisits the point x infinitely many times as T goes to infinity. A stationary process with marginal density $\pi(\cdot)$ will, for example, be recurrent with $\bar{L}(T,x)\approx T\pi(x)$ as $T\to\infty$. As we shall see, $\bar{L}(T,x)$ determines the rate of convergence of our estimator, so our estimator converges with a stochastic rate in the general case. This is in contrast to the stationary case where the convergence rate is deterministic.

We are now ready to give the precise assumptions under which we shall derive the asymptotic properties of the estimator:

Assumption 1 (existence and recurrence) *The functions μ and σ are continuously differentiable with $\sigma(\cdot)>0$, and $\{X_t\}$ is recurrent.*

Assumption 2 (boundedness) *The drift term is Lipschitz in the following sense,*

$$|\alpha(\omega_{t+\Delta},\tau)-\alpha(\omega_t,\tau)|\le C|y_{t+\Delta}(\tau)-y_t(\tau)|, \tag{15}$$

as $\Delta\to 0$ for some constant $C>0$; the drift term is bounded,

$$\frac{1}{\bar{L}(T,x)}\int_0^T K_h(X_t-x)|\alpha(\omega_t,\tau)|\,dt = O_P(1), \tag{16}$$

as $h \to 0$ uniformly over τ. The functions $\tau \mapsto \alpha(\omega_t, \tau)$ and $(x, \tau) \mapsto \gamma(x, \tau)$ are continuous. The coefficient $\kappa_{y,T}(\tau) = \max_i \sup_{t \in [i\Delta,(i+1)\Delta]} |y_t^*(\tau) - y_{t_i}^*(\tau)| \leq \bar{\kappa}_{y,T}$ a.s. with $\bar{\kappa}_{y,T} = O_P(\Delta^\alpha)$ for some $\alpha \in (0, \frac{1}{2})$.

Assumption 3 (observation errors) *The functions $(x, \tau) \mapsto m(x, \tau)$ and $(x, \tau) \mapsto \nu(x, \tau)$ are continuous. The coefficient $\kappa_{z,T}(\tau) = \max_i \sup_{t \in [i\Delta,(i+1)\Delta]} |z_t(\tau) - z_{t_i}(\tau)| \leq \bar{\kappa}_{z,T}$ a.s. with $\bar{\kappa}_{z,T} = O_P(\Delta^\alpha)$ for some $\alpha \in (0, \frac{1}{2})$. The correlation function $\rho\ (j) = O(j^{-a})$ for some $a > 1$.*

Assumption 4 (the kernel) *The kernel K has compact support ($[-C_1, C_1]$, say), is symmetric about zero, and is continuously differentiable. Moreover, $\int_{-\infty}^{+\infty} K(z)dz = 1$, $\int_{-\infty}^{+\infty} K^2(z)\, dz < \infty$, $\int_{-\infty}^{+\infty} z^j K(z)\, dz = 0$, $j = 1, \ldots, m-1$, and $\int_{-\infty}^{+\infty} z^m K(z)\, dz < \infty$ for some $m \geq 2$.*

Assumption 5 (sampling along the time dimension) *As $n \to \infty$, the sample frequency $\Delta \to 0$; $T = n\Delta \to \infty$; the bandwidth parameter $h_x \to 0$; and for some $\beta \in (\alpha, \frac{1}{2})$, $\Delta^\beta h_x^{-1} \bar{L}(T, x) = O_P(1)$.*

Assumption 6 (sampling along the maturity dimension) *$\tau_j = j/J$, $h_\tau \to 0$ and $h_\tau^2 J^a \to \infty$.*

Assumption 7 (asymptotic normality) *(i) $h_x^5 \Delta^{-1} \bar{L}(T, x) \to 0$, (ii) $J^{a-1} h_\tau^{3/2} \to \infty$, and (iii) $J^2 h_\tau^{2m+1} h_x \bar{L}(T, x) \Delta^{-1} \to 0$.*

In the following, let $\|f\|_\infty = \sup_{\tau \in [0,1]} \| f(\tau) \|$. As shown in the appendix, the estimator $\hat{\gamma}(x, \cdot)$ then has the following asymptotic properties under these assumptions

Theorem 3 *If Assumptions 1–6 hold, then $\|\hat{\gamma}(x, \tau) - \gamma(x, \tau)\|_\infty \to^P 0$. If, in addition, Assumption 7 holds,*

$$J\sqrt{\frac{h_\tau h_x \bar{L}(T, x)}{\Delta}}[\hat{\gamma}(x, \tau) - \gamma(x, \tau)] \xrightarrow{d} N\left(0, \frac{4\|K\|_2^4 \gamma^2(x, \tau) \int_0^1 \gamma^4(x, s)\, ds}{\left[\int_0^1 \gamma^2(x, s)\, ds\right]^2}\right),$$

Remarks

1. A necessary and sufficient condition for recurrence of $\{X_t\}$ can be found in Bandi and Phillips (2003).

2. Assumption 2 is a high-level one used to avoid any assumptions about the precise structure of the information variable, ω_t, and the drift term, α. If one, for example, is ready to assume that ω_t is a finite dimensional continuous semimartingale, one can leave out Equation (15). If in fact $\omega_t = X_t$, then Equation (16) can also be removed.

3. The long-span assumption in Assumption 5, $T \to \infty$, can actually be dropped for the single-factor model.

4. A consistent estimator of $L(T,x)$ is $\hat{L}(T, x) = \Delta \sum_{i=1}^{n} K_{h_x}(X_i - x)$ cf. Bandi and Phillips (2003:Theorem 1 and Corollary 1)].

Even if observation errors are not present in the observed yields, the estimator proposed here is of interest since it has better asymptotic properties compared to the $\tilde{\gamma}$-estimator of Jeffrey Linton, and Nguyen (1999a) as presented in Equation (2): While $\tilde{\gamma}$ converges with rate $\sqrt{h_x \bar{L}(T,x)\Delta^{-1}}$, $\hat{\gamma}$ converges with the faster rate $J\sqrt{h_\tau}\sqrt{h_x \bar{L}(T,x)\Delta^{-1}}$, since $J^2 h_\tau \to \infty$. This owes to the fact that we make use of all the information available across the different maturities, in contrast to Jeffrey, Linton, and Nguyen (1999a), where only information at one specific maturity point is used. That is, they only consider "large T" asymptotics, while we incorporate "large T and large J" asymptotics, in our results. The faster convergence rate carries over to the multifactor model proposed in the next section.

As a counterpart to the above result, we examine what happens in the case where our assumption about a vanishing correlation structure as $J \to \infty$ does not hold. That is, we do not assume $\rho_J(|\tau_j - \tau_k|) = \rho(|j - k|)$.

Theorem 4 *Assume that Assumptions 1–5 hold but that $\rho_J(|\tau_j - \tau_k|) = \rho(|j - k|)$ does not hold. If $\tau \mapsto \rho(\tau)$ is continuous, $\tau_j = j/J$, $h_\tau \to 0$, and $h_\tau J \to \infty$, then $\|\hat{\gamma}(x,\tau) - \gamma_*(x,\tau)\|_\infty \to^P 0$, where $\gamma_*(x,\cdot)$ is the solution to*

$$\gamma_*(x,\tau) = \frac{\int_0^1 [\gamma(x,s)\gamma(x,\tau) + \rho(|s-\tau|)\nu(x,s)\nu(x,\tau)]\gamma_*(x,s)\,ds}{\int_0^1 \gamma_*^2(x,s)ds}.$$

If, in addition, $h_x^5 \Delta^{-1} \bar{L}(T,x) \to 0$, $h_\tau^5 J \to 0$, then

$$J\sqrt{\frac{h_\tau h_x \bar{L}(T,x)}{\Delta}}[\hat{\gamma}(x,\tau) - \gamma_*(x,\tau)] \xrightarrow{d} N(0, V(x,\tau)).$$

We do not give an expression for $V(x, \tau)$ here, since this is a complicated functional of γ and ν. The above theorem states that if the correlation is very strong, our estimator will be biased since $\gamma(x,\cdot) \le \gamma_*(x,\cdot) \le \gamma(x,\cdot) + \nu(x,\tau)$. The lower inequality is only satisfied if $\rho(|s - \tau|) \equiv 0$, and the upper one only if $\rho(\tau) \equiv 1$. So in the extreme case where $\rho(\tau) \equiv 1$, $\gamma_*(x,\cdot) = \gamma(x,\cdot) + \nu(x,\cdot)$. Thus, when the measurement errors are perfectly correlated, our estimator performs just as badly as the naïve estimator $\tilde{\gamma}$ in Equation (2) in terms of bias. But in the intermediate case, the bias of our estimator will be strictly smaller than $\nu(x,\tau)$. Thus, even if the correlation is very strong, our estimator will perform better than $\tilde{\gamma}$ in terms of bias.

4 MULTIFACTOR MODELS

Multifactor models, both path-independent and path-dependent, have been widely studied in the literature in an attempt to improve the fit of the model when measured against empirical data, for instance, capturing dynamics of the observed short-term interest rate.[7] Balancing between the flexibility provided by more factors and potential overfitting problems, plus losing analytical tractability associated with more factors, researchers have commonly proposed two- or three-factor models as cited earlier.

The above approach can be extended to the general multifactor case in a straightforward fashion, at least for the step of setting up the criterion function. Derivation of the solution does become more burdensome and tedious compared to the one-factor model, as demonstrated below, but the basic idea is very much the same. However, major complications arise when deriving the asymptotic properties of the estimator. The most difficult hurdle arises from the nonexistence of local time of diffusion processes in higher dimensions, which makes the derivation of the asymptotic distribution challenging. For example, for a d-dimensional Brownian motion with $d \geq 3$, local time does not exist. For the case of $d=2$, although the local time does exist, its nonparametric estimate converges at a log (n) rate such that a large number of observations are needed for reliable inference. Some work has been done for nonparametric estimation of multivariate diffusions, however, most notable being the recent work by Bandi and Moloche (2001) [see also Brugiere (1991) and Knight, Li, and Yuan (1999)]. As demonstrated in Bandi and Moloche (2001), one is still able to estimate the drift and diffusion nonparametrically using kernel methods. In particular, they show consistency and mixed asymptotic normality under fairly general conditions. So even though the local time of the multidimensional diffusion process, $\{X_t\}$, might not exist, $\hat{L}(T,x) \equiv \Delta \sum_{i=1}^{n} \mathbf{K}_h(X_i - x)$, with $\mathbf{K}$ being a multidimensional kernel, can still be interpreted as a density estimator of a measure of the occupation time of $\{X_t\}$, and will under suitable conditions still converge when properly normed [see, e.g., Bandi and Moloche (2001:Theorem 3)]. We shall employ their results to derive the asymptotic properties of our estimator in the multifactor case.

We maintain the model of Equation (4), but now the diffusion term, $\gamma=(\gamma_1,\ldots,\gamma_M)$, is a $1 \times M$-dimensional vector function and $W_t^* = (W_{1,t}^*,\ldots, W_{M,t}^*)^\top$ is an M-dimensional standard Brownian motion. The process $\{X_t\}$ is now M-dimensional, but is still assumed to solve an SDE of the form Equation (5), where $\mu:\mathbb{R}^M \mapsto \mathbb{R}^M$, $\sigma:\mathbb{R}^M \mapsto \mathbb{R}^{M\times M}$ is a matrix function and $\{W_t\}$ is a M-dimensional Brownian motion, possibly identical to $\{W_t^*\}$. The factors are incorporated in the dynamics of the observation errors as before, with $\{W_t\ (\tau)\}$ given as in the single-factor model. We then wish to set up an estimator of γ in this more general setting.

The new state variables, $X_2,\ldots,X_M$, can include, for example, the long-term rate along the line of Brennan and Schwarz (1979), the forward rate itself, or the

[7] The practical implication of this particular limitation on pricing fixed-income assets is debatable. Failure to capture the dynamics of the short rates accurately does not necessarily impair the ability to price fixed-income assets. For instance, Buser, Hendershott, and Sanders (1990) and Hull and White (1990) claim that one-factor interest rate models with flexible specifications can generate interestrate derivatives prices similar to those of two-factor models. Nevertheless, even if their thesis is valid, the search for models that fit the empirical dynamics of the underlying consensus is still of great interest, at least for internal consistency.

spread between long- and short-term rate. As mentioned earlier, the analysis in Dai and Singleton (2000) demonstrates that almost all (Markovian) affine models examined in the literature implicitly or explicitly impose some overidentifying restrictions on an underlying "maximal" model. This deficiency is found in non-affine models and non-Markovian models as well. For instance, the multifactor model of Knight, Li, and Yuan (1999) has severe restrictions embedded in their specification. In our model we have retained generality as far as possible by allowing the volatility structure to be a function of all state variables rather than simpler specifications such as (for $M=2$) $\gamma_1(x,\tau)=\gamma_1(x_1,\tau)$ and $\gamma_2(x,\tau)=\gamma_2(x_2,\tau)$, where each of the volatility terms depends on just a single state variable. Strictly speaking, our model does not include the nonparametric two-factor model proposed by Pearson and Zhou (1999) as a special case. They restrict their model such that the randomness enters through just one Brownian motion, although the volatility is extended to a function of two variables. We can, however, easily allow for the dimension of the Brownian motions to differ from the number of factors. The estimator proposed below will still work in this case, only the asymptotic variance presented in Theorem 5 will not be valid anymore.

The conditional moment condition corresponding to the M-factor model is

$$E(\Delta y_i(s)\Delta y_i(\tau)|\mathcal{F}_{t_i}) \simeq \gamma(X_i, s)\gamma(X_i, \tau)^\top\Delta + \nu(X_i,s)\nu(X_i,\tau)\rho(|s-\tau|)\Delta.$$

As before, we incorporate the off-diagonal information based on this restriction into a localized criterion function to avoid any bias term to arise from the measurement errors. Proceeding as in the one-factor case, using the functional delta method, we obtain the following representation for the multifactor estimator

$$\hat{\gamma}(x,\tau) = \left[\int \hat{H}_2(s,\tau)\hat{\gamma}(x,s)^\top\hat{\gamma}(x,s)\,ds\right]^{-1}\int \hat{H}_1(x,s,\tau)\hat{\gamma}(x,s)\,ds, \tag{17}$$

where $\hat{H}_1(x,s,\tau)$ and $\hat{H}_2(s,\tau)$ are given as in the single-factor case, only now

$$\hat{v}(x,s,\tau) = \frac{\sum_{i=1}^n \mathbf{K}_{h_x}(X-x)\Delta y_i(s)\Delta y_i(\tau)}{\Delta\sum_{i=1}^n \mathbf{K}_{h_x}(X-x)}, \tag{18}$$

where $\mathbf{K}_{h_x}(x) \equiv \prod_{i=1}^M K_{h_x}(x_i)$. This is simply the multivariate version of the $\hat{v}$ used in the single-factor model. The estimator can be computed by iterating Equation (17) starting from some given initial condition as in the single-factor case.

To derive the asymptotic properties in the multifactor case, we need to slightly modify some of our assumptions used in the single-factor one.

Assumption 1′ (existence and recurrence) *The functions μ and σ satisfy local Lipschitz and linear growth conditions: For any $R>0$, there exists a constant $C(R)>0$ such that*

$$\|\mu(y)-\mu(z)\| + \|\sigma(y)-\sigma(z)\| \le C(R)\|y-z\|,$$
$$\|\mu(y)\| + \|\sigma(y)\| \le C(R)\|y\|,$$

for $\|y\|, \|z\| \le R$. The process $\{X_t\}$ is null Harris (or positive) recurrent.

Assumption 5′ (sampling along the time dimension) *As* $n \to \infty$, $\Delta \to 0$; $T = n\Delta \to \infty$; $\hat{L}(T,x) \xrightarrow{P} \infty$; *the* $h_x \to 0$; *and* $h_x^{-M}\sqrt{\Delta \log(1/\Delta)}\hat{L}(T,x) \to^P 0$ *at* $x \in \mathbb{R}^M$ *in the domain of* $\{X_t\}$.

Assumption 7′ (asymptotic normality) *(i)* $h_x^{M+4}\Delta^{-1}\hat{L}(T,x) \to^P 0$, *(ii)* $J^{a-1}h_\tau^{3/2} \to \infty$, *and (iii)* $J^2 h_\tau^{2m+1} h_x^M \hat{L}(T,x)\Delta^{-1} \to^P 0$.

As shown in the appendix, under these assumptions the estimator follows a mixed normal distribution:

Theorem 5 *Under Assumptions* 1′, 2–4, 5′, *and* 6, $\|\hat{\gamma}(x,\tau) - \gamma(x,\tau)\|_\infty \to^P 0$. *If in addition, Assumption 7 holds, then*

$$J\sqrt{\frac{h_\tau h_x^M \hat{L}(T,x)}{\Delta}}[\hat{\gamma}(x,\tau) - \gamma(x,\tau)] \xrightarrow{d} N(0, \Sigma(x,\tau)),$$

where

$$\Sigma(x,\tau) = 4\|K\|_2^{2M}\left[\int_0^1 \gamma(x,s)^\top \gamma(x,s)\, ds\right]^{-1} \Omega(x,\tau)\left[\int_0^1 \gamma(x,s)^\top \gamma(x,s) ds\right]^{-1}, \quad (19)$$

and $\Omega\,(x,\tau)$ *is given in Equation* (27).

Remarks

5. A sufficient condition for null Harris recurrence can be found in Bandi and Moloche (2001).
6. For simplicity, we use a common bandwidth, h_x, for all factors. The rate of convergence of the estimator is, as expected, slowed down by a factor of h_x^{M-1} relative to the one-factor case; this is the standard curse of dimensionality. There is another potential one that enters through the stochastic term $\hat{L}(T,x)$. The influence of $\hat{L}(T,x)$ in the general case is difficult to quantify, however.
7. An equivalent result to Theorem 4 can be derived in the multifactor case, but we shall refrain from doing so here.

5 EMPIRICAL IMPLEMENTATION

We apply the above techniques to estimate the volatility structure for one-factor and two-factor models from daily Center for Research in Security Prices (CRSP) bond data from January 1961 to December 1998. The first step in estimating a dynamic model of the yield curve is to extract the unobservable yield curves themselves from coupon-bearing bonds observed in the market: LMNT's kernel smoothing-based yield curve fitting, the implementation of which is described in detail in Jeffrey, Linton, and Nguyen (1999b), is our choice of yield curve extraction here.

Our choice of kernel is the commonly used gaussian $K(z) = \exp(-z^2/2)/\sqrt{2\pi}$. This kernel does not satisfy all assumptions in Assumption 4, but we do not believe this is of great relevance since many empirical studies have indicated that the actual kernel choice is of little importance. Choosing an optimal bandwidth for a nonparametric estimator is still an elusive question in the literature, there being no single scheme that is uniformly accepted, although cross-validation is a frequently used procedure [see Härdle (1990) and Pagan and Ullah (1999) for extensive discussion of the proliferation of proposed schemes]. The existing methods are designed, however, for the standard regression context; for diffusion models, especially ones that allow for nonstationary processes, rules are yet to be developed. Similar to that of LMNT, since no closed-form solutions are available, estimation of the models are computationally demanding, thus rendering cross-validation an unattractive option. So we opt instead to use a flat bandwidth (obtained by visual inspection) in our study and leave the development of other bandwidth selection methods to future research. For the diagnostic nature of our empirical work, this appears not to be of great material import. Bandwidth for the time to maturity dimension is fixed at the one-year level, while that along the interest rate dimension is fixed at 1%.

5.1 One Factor

We estimate our model based on the first-order condition of Equation (8). Although the iterative scheme suggested by Equation (9) is feasible, we found it more convenient to solve the first-order condition Equation (8) directly by a minimization routine. This finding is consistent with the implementation procedure reported in Jeffrey, Linton, and Nguyen (1999b), where a similar but somewhat simpler first-order condition of LMNT is implemented.

The nonparametric estimates for the volatility structure of the one-factor HJM model based on interest rate data from January 1961 to December 1998 is shown in Figure 1. The well-known feature of the volatility structure reported in the literature, that interest rates become more volatile when the level is high, is again observed here. For instance, volatility becomes as high as 4% when the short-term rate hits 18% (this level is only observed in the "Fed-experiment" period from 1979 to 1982). Along the maturity dimension, volatility tends to slowly increase with time to maturity when short-term rates are low, but the pattern tends to reverse itself when interest rates drift to higher ranges. Overall the shape of the volatility surface is rather consistent with what has been observed in the empirical literature [see Jeffrey, Linton, and Nguyen (1999a) for similar results].

We also experiment with different time periods, for instance, starting the data period in 1970, 1983, and 1990, respectively. The divisions are motivated by the oil shock in the early 1970s, the so-called Fed-experiment from 1979 to 1983, where interest rates were floated by the central bank, and the relatively low and stable interest rate that prevailed in the 1990s. The volatility structures estimated for these three periods are reported in Figures 2, 3 and 4, respectively. Note that to avoid biasedness at the boundary and extrapolation, the volatility structures are

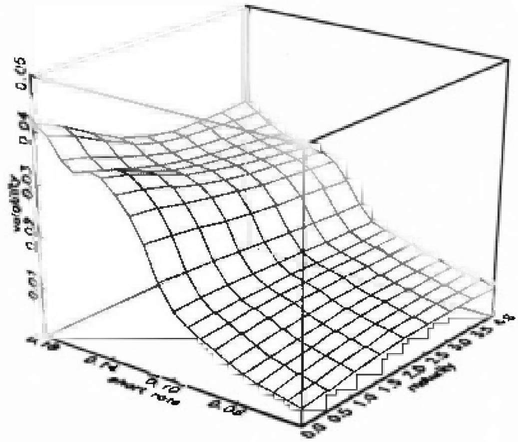

Figure 1 Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from January 1961 to December 1998. Maturity ranges from zero to four years and the short rate is from 0% to 18%.

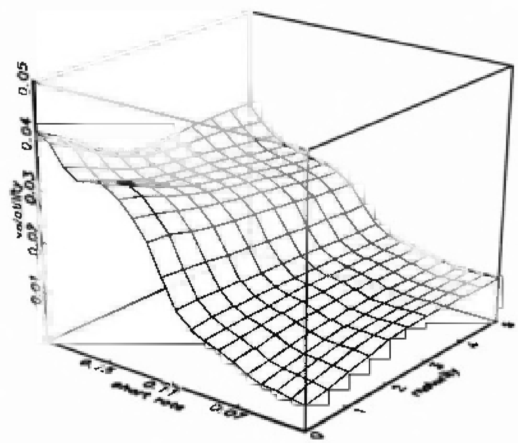

Figure 2 Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from January 1970 to December 1998. Maturity ranges from zero to five years and the short rate is from 0% to 18%.

estimated only in the ranges where the factors are observed. So in the first two figures, where the high interest rate period of 1979–1983 is included in the estimation procedures, the instantaneous interest rates (the state variable) can go from 0% to 18%, while in the last two figures they only go high up to around 10%. Similar consideration is built into the time to maturity dimension. For instance, if the data period goes back as far as 1960, we only examine the yield curves with time to maturity up to four years, since bonds with longer maturities were rarely available then. This scale increases to five and nine years if data from later periods

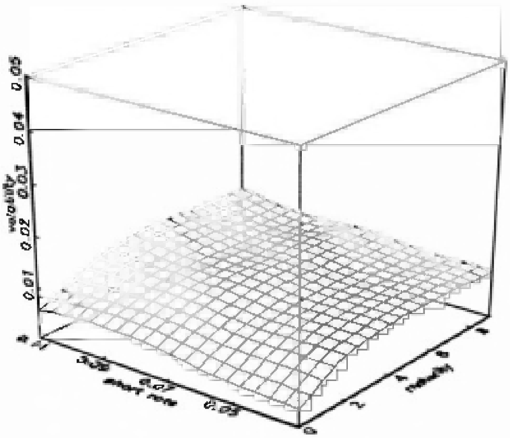

Figure 3 Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from January 1983 to December 1998. Maturity ranges from zero to nine years and the short rate is from 0% to 10%.

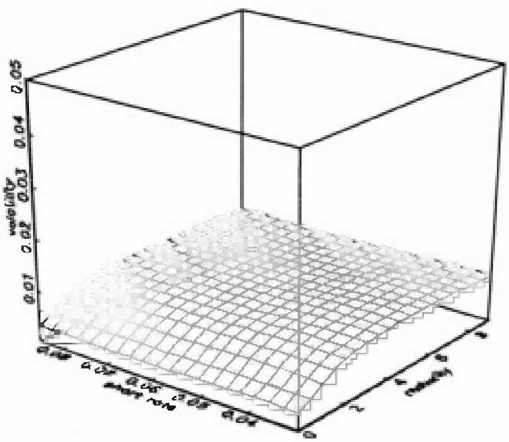

Figure 4 Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from January 1990 to December 1998. Maturity ranges from zero to nine years and the short rate is from 0% to 8%.

are used instead. For ease of comparison, however, the scale along the volatility axis remains constant across the figures.

Comparing these figures, notwithstanding the scale differences (along the axis of short rates and time to maturity), the remarkable differences between the graphs seem to suggest some nonstationary behavior for interest rates. When data from the chaotic period of the Fed-experiment is included, as in Figures 1 and 2, interest rates not only become more volatile when they are higher, but even when interest rates are low, volatility is higher compared to later periods. After the Fed-experiment period, the volatility surface has become much more stable,

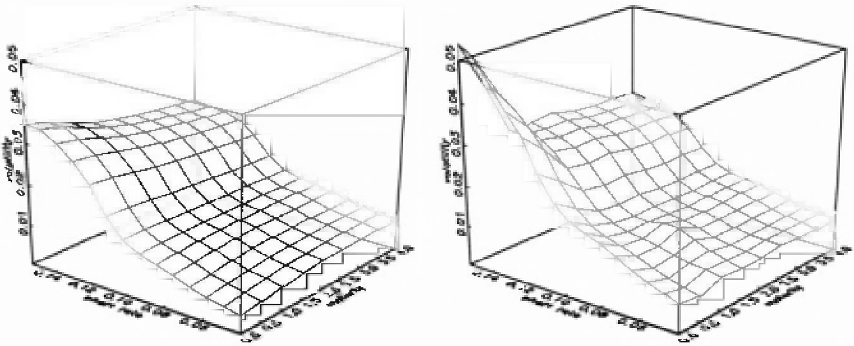

Figure 5 Volatility structure of the yield curves is estimated nonparametrically from CRSP daily bond data from January 1983 to December 1998 where the yield curve is extracted by the LMNT (2001) kernel smoothing-based method. The graph at the top is the volatility structure of the yield curves estimated nonparametrically by the method proposed in this article. The graph at the bottom is obtained by the method proposed in Jeffrey, Linton, and Nguyen (1999a). Maturity ranges from zero to four years and the short rate is from 0% to 15%.

although its typical shape is still observed; volatility increases with instantaneous interest rates and time to maturity.

In Figure 5, we compare the volatility surfaces obtained by different methodologies, that is, the one developed in this article which accounts for measurement errors, and the simpler method employed in Pearson and Zhou (1999) and Jeffrey, Linton, and Nguyen (1999a). Yield curves from 1970 to 1980 are chosen to conduct this experiment. The first graph shows the estimate of the volatility surface using the new method, which uses only off-diagonal information in the moment conditions developed above, while the second one shows that of the more "naive" method, which uses only information along the diagonal of the moment conditions. Examining the two graphs, using off-diagonal restrictions as we have in this article does not yield an estimate whose shape is dramatically different from that obtained using information from the diagonal alone. However, as expected, the latter does overestimate the volatility surface (by adding the variance of the measurement errors into its estimates). When the short rate reaches 18% for instance, its estimate of the volatility is around 0.051, while the new method yields an estimate of 0.034. This magnitude of difference is definitely material when one uses the volatility surface to price fixed-income instruments.

5.2 Two Factors

As cautioned earlier, the variance of our estimates in this case (and in models with more factors) is relatively large, making statistical inference based on these estimates difficult, besides being computationally burdensome in view of the complexity of the estimator. Consequently we cannot wholeheartedly endorse nonparametric estimation in multifactor models for inference and testing purposes. However, in the search for a reasonable parametric model, nonparametric methods can provide a good starting point even in the multivariate case.

Table 1 Volatility structure associated with the second Brownian motion from the two-factor HJM model is estimated nonparametrically from CRSP daily bond data from January 1990 to December 1998 where the yield curve is extracted by the LMNT (2001) kernel smoothing-based method.

Maturity LR (%)	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
2.51	7.54	7.78	7.97	8.12	8.25	8.38	8.46	8.48	8.41
3.34	8.52	8.79	8.98	9.14	9.31	9.46	9.52	9.47	9.35
4.18	10.11	10.36	10.53	10.71	10.94	11.13	11.15	11.02	10.85
5.01	12.47	12.57	12.58	12.65	12.85	13.05	13.05	12.87	12.67
5.84	14.84	14.70	14.43	14.25	14.31	14.43	14.42	14.24	14.07
6.67	16.58	16.23	15.69	15.23	15.06	15.07	15.02	14.88	14.75
7.51	17.82	17.27	16.51	15.81	15.42	15.30	15.22	15.11	15.03
8.34	18.87	18.13	17.14	16.22	15.65	15.41	15.30	15.21	15.16
9.17	19.91	18.92	17.67	16.52	15.79	15.46	15.32	15.26	15.22
10.01	20.92	19.63	18.06	16.66	15.78	15.39	15.24	15.19	15.17
10.84	21.79	20.17	18.26	16.60	15.60	15.16	15.02	14.99	14.98

The bandwidth chosen for the volatility structure estimation is 1%. Maturity (τ) ranges from zero to four years, short rate is fixed at 6.67%, and the long rate (LR) is from 2.51% to 10.84%. Volatility is reported in 1/1000.

For illustrative purposes, we implement a two-factor model here, with the data period from January 1970 to December 1998. As mentioned earlier, which factors to be included is commonly rather ad hoc in the literature, with a whole array of existing specifications. In our implementation, the chosen factors are the short-term and long-term interest rates, respectively, a rather common choice in the literature. The same trick used earlier is applied here, that is, solving for curves of $\gamma_1(x, \tau)$ and $\gamma_2(x, \tau)$ at a fixed point x and different τ. Even with that, the first-order condition can be difficult to solve, since we have to solve for the two curves, $\gamma_1(x, \tau)$ and $\gamma_2(x, \tau)$, simultaneously.

Graphical presentation in this case is quite problematic, since γ_1 and γ_2 are functions of three variables — the short rate, the long rate and the maturity. So for illustration we report Tables 1 and 2, which contain values of γ_2 and γ_1 respectively for one level of the short rate, $x_1 = 6.7\%$, while we allow the long-rate variable, x_2, to vary from 2.5% to 10.8% and maturities to vary from zero to four years. The ranges are chosen so that empirical data can best accommodate the estimation procedure, preempting extrapolation and boundary issues. Volatility is reported in 1/1000.

Casual observations suggest that values of γ_1 and γ_2 are rather close, and around $\eta(\cdot)\sqrt{.5}$, where η denotes the estimate of the volatility function for our one-factor model obtained in the previous section. For each volatility, a behavior similar to the one-factor case is displayed. Volatility increases with the stochastic state variable (the long rate), while along the maturity, volatility seems to

Table 2 Volatility structure associated with the second Brownian motion from the two-factor HJM model is estimated nonparametrically from CRSP daily bond data from January 1990 to December 1998 where the yield curve is extracted by the LMNT (1998) kernel smoothing-based method.

Maturity LR (%)	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
2.51	7.55	7.82	8.04	8.19	8.32	8.44	8.54	8.55	8.49
3.34	8.54	8.83	9.05	9.22	9.39	9.53	9.60	9.56	9.44
4.18	10.14	10.42	10.62	10.80	11.03	11.21	11.24	11.12	10.94
5.01	12.52	12.65	12.68	12.75	12.95	13.15	13.15	12.98	12.78
5.84	14.91	14.79	14.54	14.37	14.42	14.54	14.53	14.36	14.18
6.67	16.67	16.33	15.81	15.35	15.18	15.18	15.14	15.00	14.87
7.51	17.91	17.38	16.64	15.93	15.54	15.41	15.33	15.23	15.14
8.34	18.97	18.25	17.27	16.35	15.77	15.52	15.41	15.33	15.28
9.17	20.02	19.04	17.80	16.65	15.91	15.57	15.44	15.38	15.34
10.01	21.04	19.75	18.19	16.79	15.90	15.50	15.36	15.31	15.29
10.84	21.91	20.30	18.40	16.73	15.72	15.27	15.13	15.11	15.10

The bandwidth chosen for the volatility structure estimation is 1%. Maturity (τ) ranges from zero to four years, short rate is fixed at 6.67%, and the long rate (LR) is from 2.51% to 10.84%. Volatility is reported in 1/1000.

positively correlate with time to maturity when the long rate is low, but turns to negatively correlate with time to maturity when the long rate reaches into the higher range.

6 CONCLUSION

The HJM approach has revolutionized dynamic models of the fixed-income market, but specification and estimation issues in this framework remain a serious challenge. We have here considered a nonparametric multifactor HJM model that incorporates observation errors, and proposed an estimation method for it. The main focus has been on the single-factor model, but as demonstrated, the technique can be readily adapted to multifactor models, although the non-existence of local time in this setting complicates the derivation of the asymptotic properties of the estimator. The theoretical results indicate rather slow convergence properties for nonparametric estimators, consistent with their counterparts in nonparametric regression. But we demonstrate that the estimator proposed here has better precision than previously suggested nonparametric estimators in the literature due to the fact that we incorporate information both across maturities and over time. While we acknowledge that the nonparametric estimator presented here may not have the highest precision, we still feel that it is an attractive tool that may help researchers in finding an appropriate parametric specification.

APPENDIX A

A.1 Proof of Theorem 3

In the following we shall write $\sum_{j\neq k}$ for $\sum_{j=1}^{J}\sum_{j\neq k,k=1}^{J}$. Also observe that given that we have assumed that $\tau_j = j/J$, we may redefine our estimator as

$$\hat{\gamma}(x,\tau) = \frac{\int \hat{H}_1(x,\tau,s)\hat{\gamma}(x,s)\,ds}{\int \hat{\gamma}^2(x,s)\,ds}.$$

A.1.1 Consistency In the consistency proof, we denote the true diffusion term by γ_0. We wish to show that $\|\hat{\gamma}(x,\tau) - \gamma_0(x,\tau)\|_\infty \to^P 0$ for any given x in the domain of $\{X_t\}$. In the following let x be given. We first set up our framework. Let $C[0,1]$ denote the space of continuous functions f: $[0,1] \mapsto \mathbb{R}$. We then restrict the set of admissible functions from which we choose our estimator to

$$\Gamma = \Gamma_x = \{\gamma(x,\cdot) \in C[0,1] : \varepsilon_x \leq \gamma(x,\cdot), \|\gamma(x,\cdot)\|_\infty \leq B_x\}$$

for some $\varepsilon_x, B_x > 0$. We then choose a compact subset of Γ, which we also denote Γ; this should cause no confusion. Let

$$G_{nJ}(\gamma)(\tau) = \gamma(x,\tau)\int_0^1 \gamma^2(x,s)\,ds - \int_0^1 \hat{H}_{1,nJ}(s,\tau)\gamma(x,s)\,ds$$

for any function $\gamma\,(x,\cdot) \in \Gamma$, and define $Q_{nJ}(\gamma) = \int_0^1 |G_{nJ}(\gamma)(\tau)|\,d\tau$. We then show the desired result by checking that the conditions in Newey and Powell (2003:Lemma A2) are satisfied:

Lemma 6 *Suppose that (i) there exists some deterministic function $Q(\gamma)$ with a unique minimum at γ_0 lying in the metric space Γ; (ii) $Q_{nJ}(\gamma)$ and $Q(\gamma)$ are continuous, (iii) Γ is compact; and (iv)* $\sup_{\gamma\in\Gamma}|Q_{nJ}(\gamma) - Q(\gamma)| = o_P(1)$. *Then* $\hat{\gamma} = \arg\min_{\gamma\in\Gamma} Q_{nJ}(\gamma) \to^P \gamma_0$.

We first show that the uniform convergence holds. To prove this we introduce the partial limit,

$$\hat{H}_{1,J}(x,s,\tau) = \frac{1}{J^2}\sum_{j\neq k} K_{h_\tau}(\tau_j - \tau)K_{h_\tau}(\tau_k - s)\gamma_0(x,\tau_j)\gamma_0(x,\tau_k),$$

and the final one, $H_1\,(x,s,\tau) = \gamma_0\,(x,s)\,\gamma_0\,(x,\,\tau)$, for $s,\tau \in [0,1]$. We then split our proof into three steps: First, we show that (i) $\int_0^1\int_0^1 |\hat{H}_{1,nJ}(x,s,\tau) - \hat{H}_{1,J}(x,s,\tau)|\,ds\,d\tau \to^P 0$, $n,J \to \infty$, and then that (ii) $\int_0^1\int_0^1 |\hat{H}_{1,J}(x,s,\tau) - H_1(x,s,\tau)|\,ds\,d\tau \to 0$, $J \to \infty$. From (i) and (ii), it will easily follow that (iii) $\sup_{\gamma\in\Gamma}\int_0^1 |G_{nJ}(\gamma)(\tau) - G(\gamma)(\tau)|\,d\tau \to^P 0$, as $n,J \to \infty$, where

$$G(\gamma)(\tau) = \gamma(x,\tau)\int_0^1 \gamma^2(x,s)\,ds - \gamma_0(x,\tau)\int_0^1 \gamma_0(x,s)\gamma(x,s)\,ds.$$

Finally, by (iii) we then obtain $\sup_{\gamma\in\Gamma}|Q_{nJ}(\gamma) - Q(\gamma)| \to^P 0$, n, $J \to \infty$ with $Q(\gamma) = \int_0^1 |G(\gamma)(\tau)|\,d\tau$.

To show (i), we split up $\hat{H}_{1,nJ}(s,\tau)$ into four terms,

$$\hat{H}_{1,nJ}(x,s,\tau) = \hat{H}_{y^*y^*}(x,s,\tau) + \hat{H}_{y^*z}(x,s,\tau) + \hat{H}_{zy^*}(x,s,\tau) + \hat{H}_{zz}(x,s,\tau), \tag{20}$$

where

$$\hat{H}_{ab}(x,s,\tau) = \frac{1}{J^2}\sum_{j\neq k} K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)\hat{v}_{ab}(x,\tau_j,\tau_k),$$

$$\hat{v}_{ab}(x,s,\tau) = \frac{\sum_{i=1}^n K_{h_x}(X_i-x)\Delta a_i(s)\Delta b_i(\tau)}{\Delta\sum_{i=1}^n K_{h_x}(X_i-x)}$$

for $a,b=y^*$ and z. We then claim that $\|\hat{H}_{y^*y^*}(x,\cdot,\cdot)-H_{1,J}(x,\cdot,\cdot)\|_\infty \to^P 0$, while the three remaining terms go to zero in probability. These claims are shown in Lemmas 7–9. The claim in (ii) follows by standard techniques for kernels [see, e.g., Pagan and Ullah (1999)].

By construction, Γ is a metric compact space. It is easily seen that $\gamma \mapsto Q_{nJ}(\gamma)$ and $\gamma \mapsto Q(\gamma)$ both are continuous, while it is shown in Lemma 10 that $Q(\gamma)$ has a unique minimum at $\gamma=\gamma_0$. This proves consistency.

Lemma 7 *Under Assumptions 1–6,* $\|\hat{H}_{y^*y^*}(x,\cdot,\cdot)-H_{1,J}(x,\cdot,\cdot)\|_\infty \to^P 0$.

Proof First observe that under Assumptions 4 and 6, $J^{-2}\sum_{j\neq k}K_{h_\tau}(\tau_j-\tau)\,K_{h_\tau}(\tau_k-s)=1+O(h_\tau^m)$ uniformly over $(s,\tau)\in[0,1]\times[0,1]$ as $J\to\infty$ [see, e.g., Pagan and Ullah (1999:Theorem 2.3)]. Using this result together with the assumption that K has compact support, we obtain

$$\begin{aligned}
&\sup_{s,\tau\in[0,1]}\left|\frac{1}{J^2}\sum_{j\neq k}K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)[v_{y^*y^*}(x,\tau_j,\tau_k)-\gamma_0(x,\tau_j)\gamma_0(x,\tau_k)]\right|\\
&\quad\le \sup_{s,\tau\in[0,1]}\frac{1}{J^2}\sum_{j\neq k}K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)|v_{y^*y^*}(x,\tau_j,\tau_k)-\gamma_0(x,\tau_j)\gamma_0(x,\tau_k)|\\
&\quad\le \sup_{s,\tau\in[0,1]}|v_{y^*y^*}(x,s,\tau)-\gamma_0(x,s)\gamma_0(x,\tau)|\sup_{s,\tau\in[0,1]}\frac{1}{J^2}\sum_{j\neq k}K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)\\
&\quad= \sup_{s,\tau\in[0,1]}|v_{y^*y^*}(x,s,\tau)-\gamma_0(x,s)\gamma_0(x,\tau)|\sup_{s,\tau\in[0,1]}[1+O(h_\tau^m)].
\end{aligned}$$

The result now follows from Lemma 13. ■

Lemma 8 *Under Assumptions 1–6,* $\int_0^1\int_0^1|\hat{H}_{zz}(x,s,\tau)|\,dsd\tau \to^P 0$.

Proof First observe that, using the same arguments as in the proof of Lemma 7, $\hat{H}_{zz}(x,s,\tau)$ converges uniformly in (s,τ) toward

$$H_{zz}(x,s,\tau) = \frac{1}{J^2}\sum_{j\neq k}K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)\nu(x,\tau_j)\nu(x,\tau_k)\rho(|j-k|)$$

in probability. Next, we show that $\int_0^1\int_0^1 H_{zz}(x,s,\tau)\,dsd\tau \to 0$. As a first step, we show pointwise convergence for $|\tau - s| > h_\tau$ (where, for simplicity, we assume that the support of K is $[-1/2, 1/2]$). Since $\tau \mapsto \nu(x,\tau)$ is bounded, we may ignore this function in the following, and since K is Lipschitz and $\rho(j) = O(j^{-a})$,

$$\begin{aligned}
&\frac{1}{J^2}\sum_{j\neq k} K_{h_\tau}(\tau_j-\tau)|K_{h_\tau}(\tau_k-s)-K_{h_\tau}(\tau_j-s)|\rho(|j-k|)\\
&\quad\leq \frac{C}{h_\tau^2 J^3}\sum_{j=1}^{J} K_{h_\tau}(\tau_j-\tau)\sum_k |j-k|^{1-a}\\
&\quad\leq \frac{C}{h_\tau^2 J^2}\sum_{j=1}^{J} j^{1-a}\left\{\frac{1}{Jh_\tau}\sum_k K\left(\frac{(\tau J-j)-k}{Jh_\tau}\right)\right\}\\
&\quad\leq \frac{C}{h_\tau^2 J^2} J^{2-a}\sup_{s\in[0,1]}\left\{\frac{1}{Jh_\tau}\sum_k K_{h_\tau}(\tau_k-s)\right\}\\
&\quad= O(h_\tau^{-2}J^{-a});
\end{aligned}$$

and by

$$\left|\frac{\tau-\tau_j}{h_\tau}\right| + \left|\frac{s-\tau_j}{h_\tau}\right| \geq \left|\frac{s-\tau}{h_\tau}\right| > 1,$$

it holds that either $|\tau-\tau_j|/h_\tau > 1/2$ or $|s-\tau_j|/h_\tau > 1/2$. Since $K(x) = 0$, $|x| > 1/2$, this implies that

$$\frac{1}{J^2}\sum_{j,k} K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)|\rho(|j-k|)| = 0.$$

Second, observe that for all s, τ,

$$\begin{aligned}
|H_{zz}(x,s,\tau)| &\leq \sup_{\tau\in[0,1]}\frac{1}{J^2}\sum_{j,k} K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)\nu(x,\tau_j)\nu(x,\tau_k)\\
&= \nu(x,s)\nu(x,\tau) + O(h_\tau^m).
\end{aligned}$$

Thus, by dominated convergence,

$$\begin{aligned}
\int_0^1\int_0^1 |H_{zz}(x,s,\tau)|\,ds\,d\tau &= \int_{|s-\tau|>h_\tau} |H_{zz}(x,s,\tau)|\,dsd\tau + \int_{|s-\tau|\leq h_\tau} |H_{zz}(x,s,\tau)|\,dsd\tau\\
&= O(h_\tau^{-2}J^{-a}) + O(h_\tau^2).
\end{aligned}$$

Lemma 9 *Under Assumptions* 1–6, $\|\hat{H}_{zy}(x,\cdot,\cdot)\|_\infty \to^P 0$ *and* $\|\hat{H}_{yz}(x,\cdot,\cdot)\|_\infty \to^P 0$.

Proof This follows by the same arguments as in the proof of Lemma 7, together with the fact that $v_{z,y^*}(x,s,\tau) \to^P 0$ uniformly over $s,\tau \in [0,1]$ (cf. Lemma 13). ∎

Lemma 10 $Q(\gamma_0) < Q(\gamma)$, *for any* $\gamma \neq \gamma_0$.

Proof Define $\langle f, g\rangle = \int_0^1 f(s)g(s)\,ds$. We may then write

$$G(\gamma)(\tau) = \gamma(x,\tau)\langle\gamma,\gamma\rangle - \gamma_0(x,\tau)\langle\gamma,\gamma_0\rangle.$$

Observe that $G(\gamma)(\tau) = 0$ if and only if

$$\gamma(x,\tau) = \frac{\langle\gamma,\gamma_0\rangle}{\langle\gamma,\gamma\rangle}\gamma_0(x,\tau).$$

Thus any root of G must take the form $\gamma_a(x,\tau) = a\gamma_0(x,\tau)$, $a \neq 0$. Plugging this in,

$$\begin{aligned} G(\gamma_a)(x,\tau) &= a^3\gamma_0(x,\tau)\langle\gamma_0,\gamma_0\rangle - a\gamma_0(x,\tau)\langle\gamma_0,\gamma_0\rangle \\ &= a(a^2-1)\gamma_0(x,\tau)\langle\gamma_0,\gamma_0\rangle = 0, \end{aligned}$$

if and only if $a = \pm 1$ or 0. The negative solution and the zero solution is not feasible given the restrictions imposed on Γ. We conclude that $G(\gamma)(x,\tau) = 0$ for all $\tau \in [0,1]$ if and only if $\gamma(x,\cdot) = \gamma_0(x,\cdot)$. ■

A.1.2 Asymptotic distribution It will be convenient to use the representation in Equation (9). Using that the true volatility function satisfies $\gamma(x,\tau) = \int_0^1 H_1(x,s,\tau)\gamma(x,s)\,ds / \int_0^1 \gamma^2(x,s)\,ds$, with H_1 defined in the previous section, we make a first-order Taylor expansion of the right-hand side with respect to $\hat{\gamma}$ around γ,

$$\begin{aligned} \hat{\gamma}(x,\tau) - \gamma(x,\tau) &= \frac{\int_0^1 \hat{H}_1(x,s,\tau)\hat{\gamma}(x,s)\,ds}{\int_0^1 \hat{\gamma}^2(x,s)\,ds} - \frac{\int_0^1 H_1(x,s,\tau)\gamma(x,s)\,ds}{\int_0^1 \gamma^2(x,s)\,ds} \\ &= \frac{\int_0^1 [\hat{H}_1(x,s,\tau) - H_1(x,s,\tau)]\gamma(x,s)\,ds}{\int_0^1 \gamma^2(x,s)\,ds} \\ &\quad - \frac{\int_0^1 H_1(x,\tau,s)\gamma(x,s)\,ds \int_0^1 \gamma(x,s)[\hat{\gamma}(x,s) - \gamma(x,s)]\,ds}{[\int_0^1 \gamma^2(x,s)\,ds]^2} \\ &\quad + \frac{\int_0^1 H_1(x,s,\tau)[\hat{\gamma}(x,s) - \gamma(x,s)]\,ds}{\int_0^1 \gamma^2(x,s)\,ds} + O_P(\|\hat{H}_1 - H_1\|_2^2 + \|\hat{\gamma} - \gamma\|_2^2), \end{aligned}$$

where $\|\hat{H}_1 - H_1\|_2^2 = \int_0^1\int_0^1 [\hat{H}_1(x,s,\tau) - H_1(x,s,\tau)]^2 ds\,d\tau$, and $\|\hat{\gamma} - \gamma\|_2^2 = \int_0^1 [\hat{\gamma}(x,\tau) - \gamma(x,\tau)]^2\,d\tau$. First, going through the proof of consistency, one will realize that under Assumption 7, $\|\hat{\gamma} - \gamma\|_2^2$ and $\|\hat{H}_1 - H_1\|_2^2$ both are $O_P(M_{nJ}^{-1})$, where $M_{nJ} = h_\tau J^2 h_x \bar{L}(T,x)\Delta^{-1}$; so we can safely ignore the last term in the above expression.

In order to deal with the leading terms, we define the function $\hat{b}(x,\tau)$ by

$$\hat{b}(x,\tau) = \frac{\int_0^1 [\hat{H}_1(x,s,\tau) - H_1(x,s,\tau)]\gamma(x,s)\,ds}{\int_0^1 \gamma^2(x,s)\,ds}, \tag{22}$$

and the linear operator A by

$$A[f](x,\tau)=\frac{\int_0^1 H_1(x,s,\tau)f(x,s)\,ds}{\int_0^1\gamma^2(x,s)\,ds}-\frac{\int_0^1 H_1(x,\tau,s)\gamma(x,s)\,ds\int_0^1\gamma(x,s)f(x,s)\,ds}{[\int_0^1\gamma^2(x,s)\,ds]^2}. \quad (23)$$

We then have that

$$\hat{\gamma}(x,\tau)-\gamma(x,\tau)=A[\hat{\gamma}-\gamma](x,\tau)+\hat{b}(x,\tau),$$

or equivalently (given the inverse exists),

$$\hat{\gamma}(x,\tau)-\gamma(x,\tau)=(I-A)^{-1}[\hat{b}](x,\tau)=\hat{b}(x,\tau)+(I-A)^{-1}A[\hat{b}](x,\tau).$$

We then wish to show that $\sqrt{M_{nJ}}\hat{b}(x,\tau)$ converge in distribution while $(I-A)^{-1}A[\hat{b}](x,\tau)=o_P(M_{nJ}^{-1/2})$. This is shown in the two lemmas below.

Lemma 11 *Under Assumptions* 1–7,

$$\sqrt{M_{nJ}}\hat{b}(x,\tau)\xrightarrow{d}N\left(0,\left[4\|K\|_2^4\gamma^2(x,\tau)\int_0^1\gamma^4(x,s)\,ds\right]\Big/\left[\int_0^1\gamma^2(x,s)\,ds\right]^2\right). \quad (24)$$

Proof The desired result will hold if

$$M_{nJ}^{1/2}\int_0^1[\hat{H}_1(x,s,\tau)-H_1(x,s,\tau)]\gamma(x,s)\;ds\xrightarrow{d}N(0,4\|K\|_2^2\gamma^2\Big(x,\tau\Big)\int_0^1\gamma^4(x,s)\,ds\Big).$$

We split up $\hat{H}_1$ as in Equation (20), and then show that $M_{nJ}^{1/2}\int_0^1[\hat{H}_{y^*y^*}(x,s,\tau)-H_1(x,s,\tau)]\gamma(x,s)\,ds$ weakly converges toward a normal distribution, while the three remainder terms all are $o_P(M_{nJ}^{-1/2})$. The leading term takes the form

$$\begin{aligned}&\frac{1}{J^2}\sum_{j\neq k}\int_0^1\gamma(x,s)K_{h_\tau}(\tau_k-s)\;dsK_{h_\tau}(\tau_j-\tau)\hat{v}_{y^*y^*}(x,\tau_j,\tau_k)-\int_0^1\gamma^2(x,s)\;ds\gamma(x,\tau)\\ &\quad=\frac{1}{J^2}\sum_{j\neq k}\gamma(x,\tau_k)K_{h_\tau}(\tau_j-\tau)[\hat{v}_{y^*y^*}(x,\tau_j,\tau_k)-\gamma(x,\tau_j)\gamma(x,\tau_k)]+o_P(M_{nJ}^{-1/2}),\end{aligned}$$

By Lemma 14, the process $\hat{Z}(s,\tau)=\sqrt{h_x\bar{L}(T,x)\Delta^{-1}}\{\hat{v}_{y^*y^*}(x,s,\tau)-\gamma(x,s)\gamma(x,\tau)\}$ converges weakly toward Z (s,τ). Thus

$$\begin{aligned}&\sqrt{J^2h_\tau}\frac{1}{J^2}\sum_{j\neq k}\gamma(x,\tau_k)K_{h_\tau}(\tau_j-\tau)\hat{Z}(\tau_j,\tau_k)\\ &\xrightarrow{d}\sqrt{J^2h_\tau}\frac{1}{J^2}\sum_{j\neq k}^{\gamma}\gamma(x,\tau_k)K_{h_\tau}(\tau-\tau_j)Z(\tau_j,\tau_k)\\ &\stackrel{d}{=}N\left(0,4\|K\|_2^2\frac{h_\tau}{J^2}\sum_{j\neq k}[\gamma(x,\tau_k)K_{h_\tau}(\tau_j-\tau)]^2\gamma^2(x,\tau_j)\gamma^2(x,\tau_k)+o(1)\right)\\ &\stackrel{d}{=}N\left(0,4\|K\|_2^2\left\{\frac{h_\tau}{J}\sum_j K_{h_\tau}^2(\tau_j-\tau)\gamma^2(x,\tau_j)\right\}\left\{\frac{1}{J}\sum_k\gamma^4(x,\tau_k)\right\}+o(1)\right)\\ &\stackrel{d}{=}N\left(0,4\|K\|_2^4\gamma^2(x,\tau)\int_0^1\gamma^4(x,s)ds+o(1)\right).\end{aligned}$$

The remainder term is $O(h_\tau^m)$, which is negligible due to Assumption 7. Next, we have to show that

$$\int_0^1 \hat{H}_{z,z}(x,s,\tau)\gamma(x,s)\,ds = o_P(M_{nJ}^{-1/2}), \tag{25}$$

$$\int_0^1 \hat{H}_{y^*,z}(x,s,\tau)\gamma(x,s)\,ds = o_P(M_{nJ}^{-1/2}). \tag{26}$$

Using Equation (21) with

$$|\hat{H}_{z,z}(x,s,\tau)| \leq \frac{1}{J^2}\sum_{j\neq k} K_{h_\tau}(\tau_j-\tau)K_{h_\tau}(\tau_k-s)\rho(|j-k|)\times\left\{\sup_{s,\tau}\frac{\hat{v}_{zz}(x,s,\tau)}{\rho(|s-\tau|J)}\right\},$$

and Lemma 13, the second term is $O_P(1/\sqrt{h_x\bar{L}(T,r)\Delta^{-1}})$, while, as shown in the proof of Lemma 8, the first part is $O(h_\tau^{-2}J^{-a})$ for $|s-\tau|>h_\tau$ and $O(h_\tau^2)$ for $|s-\tau|\leq h_\tau$. In total, $M_{nJ}^{1/2}\hat{H}_{z,z}(x,s,\tau) = O_P(J^{1-a}h_\tau^{-3/2})$. The desired result now follows by Assumption 7. The proof of Equation (26) is similar. ■

Lemma 12 *Under Assumptions* 1–7, $(I-A)^{-1}A\hat{b}(x,\tau) = o_P(M_{nJ}^{-1/2})$.

Proof First we check that $(I-A)^{-1}$ actually exists: Recall $\langle f,g\rangle \equiv \int_0^1 f(s)g(s)\,ds$. By plugging $H_1(x,s,\tau)$ and $H_2(x,s,\tau)$ into Equation (23),

$$A[f](x,\tau) = \frac{\gamma(x,\tau)\langle\gamma,f\rangle}{\langle\gamma,\gamma\rangle} - \frac{2\gamma(x,\tau)\langle\gamma,\gamma\rangle\langle\gamma,f\rangle}{\langle\gamma,\gamma\rangle^2} = -\frac{\gamma(x,\tau)\langle\gamma,f\rangle}{\langle\gamma,\gamma\rangle}.$$

Thus $(I-A)f = f+\gamma(x,\tau)\ \langle\gamma,f\rangle/\langle\gamma,\gamma\rangle$, implying

$$\langle(I-A)[f],(I-A)[f]\rangle = \langle f,f\rangle + 3\frac{\langle\gamma,f\rangle^2}{\langle\gamma,\gamma\rangle} > 0.$$

Thus the inverse exists. Next, observe that as before, the second term in Equation (22) can be ignored. Thus we only need to show that

$$T_{nJ} \equiv \int_0^1\int_0^1 \hat{H}_1(x,s,\tau)\gamma(x,s)\gamma(x,\tau)\,ds d\tau = o_P(M_{nJ}^{-1/2}).$$

We have

$$\begin{aligned} T_{nJ} &= \frac{2}{J^2}\sum_{j\neq k}\left[\int_0^1 K_{h_\tau}(s-\tau_j)\gamma(x,s)\,ds\right]\left[\int_0^1 K_{h_\tau}(\tau-\tau_k)\gamma(x,\tau)\,d\tau\right]\hat{v}(x,\tau_j,\tau_k) \\ &= \frac{2}{J^2}\sum_{j\neq k}\gamma(x,\tau_j)\gamma(x,\tau_k)\hat{v}(x,\tau_j,\tau_k) + O_P(h_\tau^m), \end{aligned}$$

where the first term is $O_P\left(1/\left(J\sqrt{h_x\bar{L}(T,x)\Delta^{-1}}\right)\right)$, while $h_\tau^m\sqrt{M_{nJ}} = o_P(1)$ by Assumption 7. Finally, applying the operator $(I-A)^{-1}$ has no effect on the convergence rate. ■

A.2 Proof of Theorem 4

The "consistency" part proceeds as in the proof of Theorem 3, with the only difference being that $\hat{H}_{zz}(x,s,\tau)$ now converges toward

$$H_{zz}(x,s,\tau) = \frac{1}{J^2}\sum_{j\neq k} K_{h_\tau}(\tau_j - \tau)K_{h_\tau}(\tau_k - s)\rho(|\tau_j - \tau_k|)\nu(x,\tau_j)\nu(x,\tau_k)$$

in probability, which in turn converges toward $\rho(|\tau - s|)\nu(x,s)\nu(x,\tau)$, both uniformly in (s,τ). This shows the first part of Theorem 4.

Next, we derive the asymptotic distribution of $\hat{\gamma}(x,\tau)$. We then define $Y_t(s,\tau) = (y_t^*(s), y_t^*(\tau), z_t(s), z_t(\tau))^\top$, which solves an SDE of the form Equation (32), where $\mu_Y(\omega_t,s,\tau) = (\alpha(\omega_t,s),\ \alpha(\omega_t,\tau),\ m(\omega_t,s),\ m(\omega_t,\tau))^\top$, $B_t = (W_t^*, \tilde{W}_t\ (s), \tilde{W}_t(\tau))^\top$, $(\tilde{W}_t(s), \tilde{W}_t(\tau))^\top = \Omega^{-1/2}(W_t(s), W_t(\tau))^\top$ with Ω given as in the proof of Lemma 13, and

$$\sigma_Y(x,s,\tau) = \left[\begin{matrix}\gamma(x,s) & \gamma(x,\tau)\\ 0 & 0\end{matrix}\middle|\begin{pmatrix}\gamma(x,s) & 0\\ 0 & \nu(x,\tau)\end{pmatrix}\Omega^{1/2}\right]^\top.$$

Define $\hat{Z}(a) = \sqrt{h_x\bar{L}(T,x)\Delta^{-1}}\{\hat{v}(x,a) - v(x,a)\}$, where $a = (s,\tau), \hat{v}(x,a) = \hat{v}_{y^*y^*}(x,\ a) + 2\hat{v}_{y^*z}(x,\ a) + \hat{v}_{zz}(x,\ a)$, $v(x,a) = \gamma(x,s)\ \gamma(x,\tau + \rho(|\tau - s|)\ \nu\ (x,s)\ \nu\ (x,\tau)$. By the same arguments as in the proof of Lemma 14, $\hat{Z} \to^d Z$ on the space $C([0,1]^2,\lambda)$, where Z is a gaussian process and λ is the seminorm defined by λ $(a_1,a_2) = E[\|Z(a_1) - Z(a_2)\|^2]^{1/2}$. Once this has been obtained, we proceed as in Section A.1.2, and obtain the desired result.

A.3 Proof of Theorem 5

We shall only give a sketch of the proof since the strategy employed in the proof of the single-factor case essentially carries over to the multifactor case. We first define our estimating function $G_{nJ}(\gamma)$ as

$$G_{nJ}(\gamma) = \int_0^1 \gamma(x,s)\gamma(x,s)^\top ds\gamma(x,\tau) - \int_0^1 \hat{H}_{1,nJ}(x,s,\tau)\gamma(x,s)\ ds.$$

Thus $G_{nJ}\ (\gamma)$ takes the same form as in the single-factor case. We claim that $\hat{v}$ given in Equation (18) satisfies $\hat{v}(x,s,\tau) \xrightarrow{P} \gamma(x,s)\gamma(x,\tau)^\top$ uniformly in (s,τ), and, $\hat{Z} \to^d Z$ where $\hat{Z}(a) = \sqrt{h_x^2\hat{L}(T,x)\Delta^{-1}}(\hat{v}(x,a) - \gamma(x,s)\gamma(x,\tau)^\top)$ and $Z(a)$ is gaussian with mean zero and covariance function

$$\mathrm{cov}[Z(a_1),Z(a_2)] = 4\|K\|_2^{2M}[\gamma(x,s_1)\gamma(x,\tau_1)^\top][\gamma(x,s_2)\gamma(x,\tau_2)^\top].$$

If this holds, the desired results can be shown along the same lines as the proof of Theorem 3. But under the new set of assumptions, we can prove these claims by using very much the same arguments as in the proofs of Lemmas 13 and 14, this time appealing to results obtained in Bandi and Moloche (2001). The asymptotic distribution is driven by

$$\hat{b}(x,\tau) = \left[\int_0^1 \gamma(x,s)^\top \gamma(x,s)\ ds\right]^{-1} \int_1^0 \left[\hat{H}_1(x,s,\tau) - H_1(x,s,\tau)\right] \gamma(x,s)\ ds,$$

where $J\sqrt{h_\tau h_x^M \hat{L}(T,x)\Delta^{-1}} \int_0^1 [\hat{H}_1(x,s,\tau) - H_1(x,s,\tau)]\gamma(x,s)ds \to^d N(0, 4\|K\|_2^{2M}\Omega_J(x,\tau))$. Here, $\Omega_J(x,\tau)$ is the $M \times M$ matrix given by

$$\Omega_J(x,\tau) = \frac{h_\tau}{J^2} \sum_{j \neq k} K_{h_\tau}^2(\tau_j - \tau) \left[\gamma(x,\tau_j)\gamma(x,\tau_k)^\top\right]^2 \gamma(x,\tau_k)^\top \gamma(x,\tau_k).$$

The (q,p)th element of this matrix is

$$\begin{aligned}
&\Omega_{J,qp}(x,\tau) \\
&= \frac{h_\tau}{J^2} \sum_{j \neq k} K_{h_\tau}^2(\tau_j - \tau) \left[\gamma(x,\tau_j)\gamma(x,\tau_k)^\top\right]^2 \gamma_q(x,\tau_k)\gamma_p(x,\tau_k) \\
&= \sum_{l,m=1}^{M} \frac{h_\tau}{J^2} \sum_{j \neq k} K_{h_\tau}^2(\tau_j - \tau)\gamma_l(x,\tau_j)\gamma_l(x,\tau_k)\gamma_m(x,\tau_j)\gamma_m(x,\tau_k)\gamma_q(x,\tau_k)\gamma_p(x,\tau_k) \\
&= \|K\|_2^2 \sum_{l,m=1}^{M} \gamma_l(x,\tau)\gamma_m(x,\tau) \int_0^1 \gamma_l(x,s)\gamma_m(x,s)\gamma_q(x,s)\gamma_p(x,s)\, ds + o(1) \\
&= \Omega_{qp}(x,\tau) + o(1).
\end{aligned} \tag{27}$$

We now obtain the variance given in Equation (19).

A.4 Auxiliary Lemmas

The following two lemmas supply us with uniform consistency and asymptotic normality of the leading terms in $\hat{H}_1$. The proofs of the lemmas very much follow the steps in Bandi and Phillips (2003: Proofs of Theorems 4 and 5), and we therefore do not give full details here.

Lemma 13 *Under Assumptions* 1–6,

$$\sup_{s,\tau \in [0,1]} |\hat{v}_{y^*,y^*}(x,s,\tau) - \gamma(x,s)\gamma(x,\tau)| = o_P(1), \tag{28}$$

$$\sup_{s,\tau \in [0,1]} |\hat{v}_{z,y^*}(x,s,\tau)| = o_P(1), \tag{29}$$

$$\sup_{s,\tau \in [0,1]} |\hat{v}_{z,z}(x,s,\tau) - \rho(|s-\tau|J)\nu(x,s)\nu(x,\tau)| = o_P(1), \tag{30}$$

uniformly in $(s,\tau) \in [0,1] \times [0,1]$.

Lemma 14 *Under Assumptions* 1–7,

(i) *The process* $\{\hat{Z}(a)\}$ *given by*

$$\hat{Z}(a) = \sqrt{h_x \bar{L}(T,x)\Delta^{-1}}(\hat{v}_{y^*,y^*}(x,a) - \gamma(x,s)\gamma(x,\tau)),$$

where $a = (s,\tau)$, *converges weakly toward the gaussian process* $\{Z(a)\}$ *characterized by*

$$E[Z(a)] = 0,\ \ \mathrm{cov}[Z(a_1), Z(a_2)] = 4\|K\|_2^2\gamma(x,s_1)\gamma(x,\tau_1)\gamma(x,s_2)\gamma(x,\tau_2),$$

on the space $C([0,1]^2,\lambda)$ *of continuous functions equipped with the seminorm*

$$\lambda(a_1,a_2) = E[\|Z(a_1) - Z(a_2)\|^2]^{1/2}. \tag{31}$$

(ii) *Equations* (29) *and* (30) *hold with rate* $o_P(1/\sqrt{h_r\bar{L}(T,x)\Delta^{-1}})$.

Proof of Lemma 13 We first prove Equation (28). For given $s,\tau \in [0,1]$, define $Y_t = (y_t^*(s), y_t^*(\tau))^\top$. We have

$$Y_t - Y_{i\Delta} = \int_{i\Delta}^t \mu_Y(\omega_u)\,du + \int_{i\Delta}^t \sigma_Y(X_u)\,dB_u, \tag{32}$$

where $\mu_Y(\omega_t) = (\alpha(\omega_t,s),\alpha(\omega_t,\tau))^\top$, $\sigma_y(x) = (\gamma(x,s),\gamma(x,\tau))^\top$, and $B_t = W_t^*$. Applying the multivariate version of *Itô's* formula [see, e.g., Lipster and Shiryaev (2001): Theorem 4.5)] on Y_t,

$$\begin{aligned}\Delta y_i^*(s)\Delta y_i^*(\tau) = &\int_{i\Delta}^{(i+1)\Delta}[y_t^*(\tau) - y_i^*(\tau)]\alpha(\omega_t,s)\,dt + \int_{i\Delta}^{(i+1)\Delta}[y_t^*(s) - y_i^*(s)]\alpha(\omega_t,\tau)\,dt\\ &+\int_{i\Delta}^{(i+1)\Delta}[y_t^*(\tau) - y_i^*(\tau)]\gamma(X_t,s)\,dW_t^*\\ &+\int_{i\Delta}^{(i+1)\Delta}[y_t^*(s) - y_i^*(s)]\gamma(X_t,\tau)\,dW_t^* + \int_{i\Delta}^{(i+1)\Delta}\gamma(X_t,s)\gamma(X_t,\tau)\,dt,\end{aligned}$$

we are able to write

$$\hat{v}_{y^*,y^*}(x,s,\tau) = A(s,\tau) + \sum_{i=1}^{2} B_i(s,\tau) + \sum_{i=1}^{2} V_i(s,\tau), \tag{33}$$

where

$$A(s,\tau) = \frac{\frac{1}{\Delta}\sum_{i=1}^n K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta}\gamma(X_t,s)\gamma(X_t,\tau)\,dt}{\sum_{i=1}^n K_{h_x}(X_i - x)}$$

$$B_1(s,\tau) = \frac{\frac{1}{\Delta}\sum_{i=1}^n K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta}[y_t^*(\tau) - y_i^*(\tau)]\alpha(\omega_t,s)\,dt}{\sum_{i=1}^n K_{h_x}(X_i - x)},$$

$$B_2(s,\tau) = \frac{\frac{1}{\Delta}\sum_{i=1}^{n} K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta}[y_t^*(s) - y_i^*(s)]\alpha(\omega_t,\tau)\,dt}{\sum_{i=1}^{n} K_{h_x}(X_i - x)}$$

$$V_1(s,\tau) = \frac{\frac{1}{\Delta}\sum_{i=1}^{n} K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta}[y_t^*(\tau) - y_i^*(\tau)]\gamma(X_t,s)\,dW_t^*}{\sum_{i=1}^{n} K_{h_x}(X_i - x)}$$

$$V_2(s,\tau) = \frac{\frac{1}{\Delta}\sum_{i=1}^{n} K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta}[y_t^*(s) - y_i^*(s)]\gamma(X_t,\tau)\,dW_t^*}{\sum_{i=1}^{n} K_{h_x}(X_i - x)}$$

We claim that $A(s,\tau) \to^P \gamma(x,s)\ \gamma(x,\tau)$ while the remaining terms all converge to zero uniformly. We start out by showing pointwise convergence. The proof of this follows along the same lines as in Bandi and Phillips (2003). First, we realize that

$$A(s,\tau) = \frac{\int_0^T K_{h_x}(X_t - x)\gamma(X_t,s)\gamma(X_t,\tau)\,dt + o_P(1)}{\int_0^T K_{h_x}(X_t - x)\,dt + o_P(1)}. \tag{34}$$

To show this, define $\kappa_{x,n} = \max_i \sup_{t\in[i\Delta,(i+1)\Delta]}|X_t - X_i|$; we have $\kappa_{x,n} = O_P(\Delta^\alpha)$ for all $\alpha \in (0,1/2)$ [cf. Bandi and Phillips (2003)]. Furthermore, by Assumption (6), there exists a function $D(u,\varepsilon)$ such that $|K(u) - K(v)| \le D(u,\varepsilon)\ |u - v|$ and $\lim_{\varepsilon\downarrow 0}\int_{-\infty}^{+\infty} D(u,\varepsilon)du < \infty$. Thus

$$\begin{aligned}
&\left|\sum_{i=1}^{n} K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta}\gamma(X_t,s)\gamma(X_t,\tau)\ dt - \int_0^T K_{h_x}(X_t - x)\gamma(X_t,s)\gamma(X_t,\tau)\,dt\right|\\
&\le \sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta}|K_{h_x}(X_i - x) - K_{h_x}(X_t - x)|\gamma(X_t,s)\gamma(X_t,\tau)\,dt\\
&\le \frac{\kappa_{r,n}}{h_x}\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta} D\left(\frac{X_t - x}{h_x}, \frac{\kappa_{x,n}}{h_x})\gamma(X_t,s\right)\gamma(X_t,\tau)\,dt\\
&= \frac{\kappa_{x,n}}{h_x}\int_0^T D\left(\frac{X_t - x}{h_x}, \frac{\kappa_{x,n}}{h_x}\right)\gamma(X_t,s)\gamma(X_t,\tau)\,dt\\
&\le \frac{\kappa_{x,n}}{h_x}\int_0^T D\left(\frac{X_t - x}{h_x}, \frac{\kappa_{x,n}}{h_x}\right)\bar{\gamma}^2(X_t)\,dt,
\end{aligned}$$

where $\bar{\gamma}(x) = \sup_{\tau\in[0,1]}\gamma(x,\tau)$, which is well-defined and continuous by Assumption (2). By Assumption (5), $\kappa_{x,n}/h_x = o_P(1)$, which together with Equation (14) implies

$$\begin{aligned}
\int_0^T D\left(\frac{X_t - x}{h_x}, \frac{\kappa_{x,n}}{h_x}\right)\bar{\gamma}^2(X_t)\,dt &= \frac{1}{h_x}\int_{-\infty}^{+\infty} D\left(\frac{a - x}{h_x}, \frac{\kappa_{x,n}}{h_x}\right)\bar{\gamma}^2(a)\bar{L}(T,a)\,da\\
&= \int_{-\infty}^{+\infty} D\left(q, \frac{\kappa_{x,n}}{h_x}\right)\bar{\gamma}^2(x + hq)\bar{L}(T, x + hq)\,dq\\
&= \bar{L}(T,x)\bar{\gamma}^2(x) + o_P(1).
\end{aligned}$$

This establishes Equation (34). Similarly we may prove that $\sum_{i=1}^{n} K_{h_x}(X_i - x) = \int_0^T K_{h_x}(X_t - x)\,dt + o_P(1)$. Next, applying the Darling-Kac theorem [cf. Revuz and Yor (1991:Theorem 3.12)] along the same lines as in Bandi and

Phillips (2003), it follows that

$$\frac{\int_0^T K_{h_x}(X_t - x)\gamma(X_t, s)\gamma_0(X_t, \tau)\, dt}{\int_0^T K_{h_x}(X_t - x)\, dt} = \frac{\gamma(x, s)\gamma(x, \tau)s(x) + o_P(1)}{s(x) + o_P(1)},$$

where $s(\cdot)$ is the speed function of $\{X_t\}$. Since $[0,1]^2$ is compact and we have pointwise convergence, the uniform convergence will follow if we can establish stochastic equicontinuity [see, e.g., Newey (1991)]. But for any $d, \varepsilon > 0$, there exists $\delta > 0$ such that

$$\sup_{\{v:|x-v|\le d\}} \sup_{\{s,s',\tau,\tau' || s-s'|,|\tau-\tau'|\le\varepsilon\}} |\gamma(v,s)\gamma(v,\tau) - \gamma(v,s')\gamma(v,\tau')| \le \delta.$$

This follows from the continuity of γ_0 and the fact that $\{v: |x-v| \le d\} \times [0,1]^2$ is compact. Also, observe that the discretization bias in Equation (34) is uniform over (s,τ). Thus

$$\begin{aligned}
&\sup_{|s-s'|,|\tau-\tau'|\le\varepsilon} |A(s,\tau) - A(s',\tau')| \\
&\le \sup_{|s-s'|,|\tau-\tau'|\le\varepsilon} \frac{\int_0^T K_{h_x}(X_t - x)|\gamma(X_t,s)\gamma(X_t,\tau) - \gamma(X_t,s')\gamma(X_t,\tau')|\, dt}{\int_0^T K_{h_x}(X_t - x)} + o_P(1) \\
&\le \frac{\int_0^T K_{h_x}(X_t - x)}{\int_0^T K_{h_x}(X_t - x)}\delta + o_P(1) \\
&= \delta + o_P(1),
\end{aligned}$$

where we have used the compactness of K. We have now established that the first term in Equation (33) converges uniformly. For the remaining terms, a similar strategy can be applied. We shall only consider the first one,

$$\begin{aligned}
|B_1(s,\tau)| &\le \frac{\frac{1}{\Delta}\sum_{i=1}^n K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta} |y_t(\tau) - y_i(\tau)||\alpha(\omega_t, s)|\, dt}{\sum_{i=1}^n K_{h_x}(X_i - x)} \\
&\le \frac{\bar{\kappa}_{y,T}\frac{1}{\Delta}\sum_{i=1}^n K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta} |\bar{\alpha}(\omega_t)| dt}{\sum_{i=1}^n K_{h_x}(X_i - x)},
\end{aligned}$$

where $\bar{\alpha}(\omega_t) = \sup_{\tau\in[0,1]} \alpha(\omega_t, \tau)$. From Equation (16) we obtain

$$\sum_{i=1}^n K_{h_x}(X_i - x)\int_{i\Delta}^{(i+1)\Delta} |\bar{\alpha}(\omega_t)|\, dt = \int_0^T K_{h_x}(X_t - x)|\bar{\alpha}(\omega_t)|\, dt + o_P(1) = O_P(\hat{L}(T,x)),$$

uniformly in s, while $\bar{\kappa}_{y,T} = o_P(1)$ since $\Delta \to 0$. This shows pointwise convergence; uniform convergence is now established in the same way as for the first term.

The claims of Equations (29) and (30) follow along the same lines. For example, defining B_t by

$$\begin{pmatrix} B_{1,t} \\ B_{2,t} \end{pmatrix} = \Omega^{-1/2}\begin{pmatrix} W_t(s) \\ W_t(\tau) \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & \rho_J(|s-\tau|) \\ \rho_J(|s-\tau|) & 1 \end{pmatrix},$$

yields a two-dimensional Brownian motion. Thus $Y_t = (z_t(s),\ z_t(\tau))^\top$ solves an SDE of the form of Equation (32) with $\mu_Y(\omega_t) = (m(\omega_t, s), m(\omega_t, \tau))^\top$, and $\sigma_Y(x) = \mathrm{diag}(\nu(x, s),\ \nu(x,\tau))\Omega^{1/2}$. By Itô's formula,

$$\begin{aligned}\Delta z_i(s)\Delta z_i(\tau) = &\int_{i\Delta}^{(i+1)\Delta} [z_t(\tau) - z_i(\tau)]m(\omega_t, s)\,dt + \int_{i\Delta}^{(i+1)\Delta} [z_t(s) - z_i(s)]m(\omega_t, \tau)\,dt \\ &+ \int_{i\Delta}^{(i+1)\Delta} [z_t(\tau) - z_i(\tau)]\sum_{j=1}^{2} \sigma_{Y,2j}(X_t)\,dB_{j,t} \\ &+ \int_{i\Delta}^{(i+1)\Delta} [z_t(s) - z_i(s)]\sum_{j=1}^{2} \sigma_{Y,1j}(X_t)\,dB_{j,t} \\ &+ \int_{i\Delta}^{(i+1)\Delta} \nu(X_t, s)\nu(X_t, \tau)\rho_J(|s-\tau|)\,dt.\end{aligned}$$

The proof of Equation (30) now proceeds as for Equation (28).

Proof of Lemma 14 We first show (i). By Van der Vaart and Wellner (1996: Example 1.5.10), $\hat{Z} \xrightarrow{d} Z$ if (a) $(\hat{Z}(a_1), \ldots, \hat{Z}(a_k)) \to^d (Z(a_1), \ldots, Z(a_k))$ for any $k \geq 1$ and any $k \geq 1$, where $a_i = (s_i, \tau_i) \in [0,1]^2$, $i = 1, \ldots, k$; (b) $\hat{Z}$ is stochastically equicontinuous w.r.t. λ; and (c) $[0,1]^2$ is totally bounded w.r.t. λ. Condition (b) was shown to hold in the proof of Lemma 13, and (c) clearly holds given the continuity of $\tau \mapsto \gamma(x, \tau)$. What remains to be shown is (a), which we only give a proof of for $k = 2$; this proof is easily extended to the general case. We have that $\hat{v}_{y^*,y^*}(x, a)$ can be written as in Equation (33), where we already know that the first term in Equation (33) converges toward $\gamma(x, s)\gamma(x, \tau)$; one can convince oneself that this happens with rate $o_P(1/\sqrt{h_x L(T,x)\Delta^{-1}})$. Similarly $B_i(s, \tau) = o_P(1/\sqrt{h_x L(T,x)\Delta^{-1}})$, $i = 1, 2$. What remains to be proved is that $V = (\sum_{i=1}^{2} V_i(a_1),\ \sum_{i=1}^{2} V_i(a_2))^\top$ weakly converges. The proof of this is based on embedding V in a time-changed Brownian motion. We first realize that the denominator of $\sum_{i=1}^{2} V_i(a)$ equals $M_1^{[T]}(a) + o_P(1/\sqrt{h_x L(T,x)\Delta^{-1}})$, where

$$M_u^{[T]}(a) = 2\int_0^{[uT]} K_{h_x}(X_t - x)\gamma(X_t,\ s)\gamma(X_t,\ \tau)dW_t^*. \tag{35}$$

The process $\{\sqrt{h_x}M_u^{[T]} = \sqrt{h_x}[M_u^{[T]}(a_1), M_u^{[T]}(a_2)]\}$ is a martingale with quadratic variation

$$\left[\sqrt{h_x}M^{[T]}\right]_u = \frac{4}{h_x}\int_0^{[uT]} K^2\left(\frac{X_t - x}{h_x}\right)V(X_t,\ a_1,\ a_2)\,dt \xrightarrow{a.s.} 4\|K\|_2^2 V(x,\ a_1,\ a_2)\bar{L}(uT, x),$$

where

$$V(x, a_1, a_2) = \begin{bmatrix} \gamma^2(x, s_1)\gamma^2(x, \tau_1) & \gamma(x, s_1)\gamma(x, \tau_1)\gamma(x, s_2)\gamma(x, \tau_2) \\ \gamma(x, s_1)\gamma(x, \tau_1)\gamma(x, s_2)\gamma(x, \tau_2) & \gamma^2(x, s_2)\gamma^2(x, \tau_2) \end{bmatrix}. \tag{36}$$

Let $\{\tilde{B}_u\}$ denote the Dambis, Dubins-Schwarz Brownian motion of $\{\sqrt{h_x}M_u^{[T]}\}$ [cf. Revuz and Yor (1991:Theorem 1.6)]; we then obtain

$$\sqrt{h_x}M_u^{[T]} = \tilde{B}_u \xrightarrow{d} N(0, 4\|K\|_2^2 V(x, a_1, a_2)\bar{L}(T, x)).$$

We conclude that

$$[\hat{Z}(a_1), \hat{Z}(a_2)] = \frac{\sqrt{h_x}M_1^{[T]}}{\sqrt{\bar{L}(T,x)}}\frac{\bar{L}(T,x)}{\hat{L}(T,x)} + o_P(1) \xrightarrow{d} [Z(a_1), Z(a_2)].$$

This establishes weak convergence of the marginals.

Going through the proof of pointwise convergence of Equations (29) and (30), one realises that under the additional restrictions in Assumption (7), the pointwise convergence will happen with rate $o_P(\sqrt{h_x\bar{L}(T,x)\Delta^{-1}})$. Using the same arguments as before, this can then be strengthened to uniform convergence. ■

Received August 29, 2001; revised October 17, 2002; accepted January 12, 2004

REFERENCES

Aït-Sahalia, Y. (1996). "Nonparametric Pricing of Interest Rate Derivative Securities." *Econometrica 64*, 527–560.

Balduzzi, P., S. Das, S. Foresi, and R. Sundaram. (1996). "A Simple Approach to Three-Factor Affine Term Structure Models." *Journal of Fixed Income 6*, 43–53.

Bandi, F., and G. Moloche. (2001). "On the Functional Estimation of Multivariate Diffusion Processes." Working paper, MIT.

Bandi, F., and P. C. B. Phillips. (2003). "Fully Nonparametric Estimation of Scalar Diffusion Models." *Econometrica 71*, 241–283.

Bliss, R., and P. Ritchken. (1996). "Empirical Test of Two State-Variable Heath-Jarrow-Morton Model." *Journal of Money, Credit and Banking 28*, 452–481.

Boudoukh, J., M. Richardson, R. Stanton, and R. Whitelaw. (1997). "Pricing Mortgage-Backed Securities in a Multifactor Interest Rate Environment: A Multivariate Density Estimation Approach." *Review of Financial Studies 10*, 405–446.

Breiman, L. (1991). "The $\prod$-Method for Estimating Multivariate Functions from Noisy Data (with discussion)." *Technometrics 33*, 125–158.

Brennan, M., and E. Schwartz. (1979). "A Continuous Time Approach to the Pricing of Bonds." *Journal of Banking and Finance 3*, 133–155.

Brugiere, P. (1991). "Estimation of the Coefficient of a Diffusion Process in the Multidimensional Case." *Comptes Rendus l'Academie des Sciences 312* (Serie I), 999–1004.

Bühler, W., M. Uhrig-Homburg, U. Walter, and T. Weber. (1999). "An Empirical Comparison of Forward and Spot-Rate Models for Valuing Interest Rate Option." *Journal of Finance 54*, 269–305.

Buser, S. A., P. H. Hendershott, and A. B. Sanders. (1990). "Determinants of the Value of Call Options of Default-Free Bonds." *Journal of Business 63*, S33–S50.

Carrasco, M., J. P. Florens, and E. Renault. (2003), "Linear Inverse Problems in Structural Econometrics." In J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 6. Amsterdam: North-Holland.

Chen, R. R., and L. Scott. (1993). "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates." *Journal of Fixed Income* 3, 14–31.

Chen, R. R., and L. Scott. (1992). "Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Models of the Short-Term Interest Rates." *Review of Financial Studies* 5, 613–636.

Cox, J., J. Ingersoll, and S. Ross. (1985). "A Theory of the Term Structure of Interest Rates." *Econometrica* 53, 385–406.

Cressie, N. A. (1993). *Statistics for Spatial Data*, 2nd ed. New York: John Wiley & Sons.

Dai, Q., and K. Singleton. (2000). "Specification Analysis of Affine Term Structure Models." *Journal of Finance* 55, 1943–1978.

Darolles, S., J.-P. Florens, and E. Renault. (2001). "Nonparametric Instrumental Regression." Working paper, INSEE.

de Jong, J., and P. Santa-Clara. (1999). "The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables." *Journal of Financial and Quantitative Analysis* 34, 131–157.

Duffie, D., and R. Kan. (1996). "A Yield Factor Model of Interest Rate." *Mathematical Finance* 6, 379–406.

Duffie, D., and K. Singleton. (1997). "An Econometric Model of the Term Structure of Interest Rate Swap Yields." *Journal of Finance* 52, 1287–1321.

Fama, E. F., and R. R. Bliss. (1987). "The Information in Long-Maturity Forward Rates." *American Economic Review* 77, 680–692.

Florens-Zmirou, D. (1993). "On Estimating the Diffusion Coefficient from Discrete Observations." *Journal of Applied Probability* 30, 790–804.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Hastie, T. J., and R. J. Tibshirani. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Heath, D., R. Jarrow, and A. Morton. (1992). "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation." *Econometrica* 60, 77–105.

Ho, T. S. Y., and S. B. Lee. (1986). "Term Structure Movements and Pricing Interest Rate Contingent Claims." *Journal of Finance* 41, 1011–1029.

Hull, J., and A. White. (1990). "Pricing Interest Rate Derivative Securities." *Review of Financial Studies* 3, 573–592.

Inui, K., and M. Kijima. (1998). "A Markovian Framework in Multi-Factor Heath-Jarrow-Morton Models." *Journal of Financial and Quantitative Analysis* 33, 423–440.

Jeffrey, A., O. Linton, and T. Nguyen. (1999b). "Nonparametric Estimation of Heath-Jarrow-Morton models and a Test for Path Independence." Working paper, Yale University.

Jeffrey, A., O. Linton, and T. Nguyen. (1999a). "Term Structure Estimation and a Proxy for the Instantaneous Spot Interest Rate." Working paper, Yale University.

Jiang, G., and J. Knight. (1997). "A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short Term Interest Rate Model." *Econometric Theory* 13, 615–645.

Kantorovich, L. V., and G. P. Akilov. (1982). *Functional Analysis*, 2nd rev. ed. Oxford: Pergamon Press.

Kennedy, D. P. (1994). "The Term Structure of Interest Rates as a Gaussian Random Field." *Mathematical Finance* 4, 247–258.

Kennedy, D. P. (1997). "Characterizing Gaussian Models of the Term Structure of Interest Rates." *Mathematical Finance* 7, 107–18.

Knight, J., F. Li, and M. Yuan. (1999). "Pricing Interest Rate Derivatives in a Non-Parametric Two-Factor Term Structure Model." Working Paper 99-19, Bank of Canada.

Linton, O., E. Mammen, J. P. Nielsen, and C. Tanggaard. (2001). "Estimating the Yield Curve by Kernel Smoothing Methods." *Journal of Econometrics* 105, 185–223.

Liptser, R., and A. N. Shiryaev. (2001). *Statistics of Random Processes, Vol. 1: General Theory*. New York: Springer-Verlag.

Litterman, R., and J. Scheinkman. (1991). "Common Factors Affecting Bond Returns." *Journal of Fixed Income* 1, 54–61.

Longstaff, F. A., and E. S. Schwartz. (1992). "Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model." *Journal of Finance* 47, 1259–1282.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. New York: John Wiley & Sons.

McCulloch, J. H. (1971). "Measuring the Term Structure of Interest Rates." *Journal of Business* 44, 19–31.

McCulloch, J. H. (1975). "The Tax-Adjusted Yield Curve." *Journal of Finance* 30, 811–830.

Mammen, E., O. Linton, and J. P. Nielsen. (1999). "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm Under Weak Conditions." *Annals of Statistics* 27, 1443–1490.

Newey, W. K. (1991). "Uniform Convergence in Probability and Stochastic Equicontinuity." *Econometrica* 59, 1161–1167.

Newey, W. K., and J. L. Powell. (2003). "Instrumental Variables Estimation of Nonparametric Models." *Econometrica* 71, 1565–1578.

Opsomer, J. D., Y. Wang, and Y. Yang. (2001). "Nonparametric Regression with Correlated Errors." *Statistical Science* 16, 134–153.

Pagan, A., and A. Ullah. (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Pearson, N., and T. Sun. (1994). "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model." *Journal of Finance* 49, 1279–1304.

Pearson, N. D., and A. Zhou. (1999). "A Nonparametric Analysis of the Forward Rate Volatilities." Working paper OFOR = 9905, The Office for Futures and Options Research, University of Illinois at Urbana-Champaign.

Phillips, P. C. B., and J. Park. (1998). "Nonstationary Density Estimation and Kernel Autoregression." Cowles Foundation Discussion Paper 1181, Yale University.

Revuz, D., and M. Yor. (1991). *Continuous Martingales and Brownian Motion*. New York: Springer-Verlag.

Robinson, P. M. (1997). "Large Sample Inference for Nonparametric Regression with Dependent Errors." *Annals of Statistics* 25, 2054–2083.

Santa-Clara, P., and D. Sornette. (2001). "The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks." *Review of Financial Studies* 14, 149–185.

Schaefer, S. M., and E. S. Schwartz. (1984). "A Two Factor Model of the Term Structure: An Approximate Analytical Solution." *Journal of Financial and Quantitative Analysis* *19*, 413–424.

Stanton, R. (1997). "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk." *Journal of Finance* *52*, 1973–2002.

Van der Vaart, A. W., and J. A. Wellner. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Vasicek, O. (1977). "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics* *5*, 177–88.